

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 8, 2003, 05:00:24 ; Search time 2731 Seconds
(without alignments)
3666.585 Million cell updates/sec

Title: US-09-889-756A-2
Perfect score: 2019
Sequence: 1 MAFYAFAMRAALAAVA.....AAGSGVQTASEAKTASAE 412

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 1215238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2.1/USPTO.spool/US09889756/runat.05092003.124325.23309/app.query.fasta_1.583
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=DIOSUM62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=DCT -THR MAX=100 -THR MIN=0 -ALIGN=40 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09889756 @CGN 1.1 2810 @runat.05092003.124325.23309 -NCPU=6 -ICPU=3
-NO MAP -LARGEOUTRY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estma:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vit:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vit:*
28: gb_gss1:*

29: gb_gss2:*
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	585	29.0	893	29	CNS01JMC
2	512.5	25.4	834	29	CNS01JPC
3	411.5	20.4	642	28	BH368028 Ag-ND-138
4	370.5	18.4	601	28	BH190105
5	343	17.0	1096	29*	BZ558375
6	318.5	15.8	1503	29	BZ569193
7	315	15.6	775	29	BZ550352
8	304	15.1	515	28	BH190022
9	291.5	14.4	1236	29	BZ554554
10	280	13.9	798	28	BH250378
11	269.5	13.3	620	28	BH189811
12	261.5	13.0	1263	29	BZ557583
13	254	12.6	937	29	BZ562652
14	250	12.4	1417	29	BZ574393
15	249	12.3	802	10	BF663822
16	246	12.2	620	28	BH190195
17	244	12.1	596	28	BH189826
18	241.5	12.0	1108	29	BZ577818
19	233	11.5	1259	29	BZ580098
20	220	10.9	470	29	BZ332913
21	219.5	10.9	923	29	BZ554963
22	216	10.7	597	28	BH189931
23	216	10.7	747	29	BZ561644
24	211.5	10.5	568	29	BZ336018
25	210.5	10.4	772	29	BZ579248
26	194	9.6	809	29	BZ551287
27	189.5	9.4	595	14	CA554488
28	189.5	9.4	766	29	BZ558130
29	188	9.3	775	28	AZ933876
30	186	9.2	607	29	BZ340887
31	186	9.2	1379	29	BZ569194
32	180.5	8.9	472	28	AZ302989
33	179.5	8.9	450	28	BH247218
34	177	8.8	745	28	BH242325
35	173.5	8.6	847	29	BZ562720
36	172.5	8.5	793	29	BZ549534
37	170.5	8.4	876	29	BZ638338
38	168.5	8.3	469	10	BF416737
39	168	8.3	1307	29	BZ574481
40	163.5	8.1	1228	29	BZ578763
41	163	8.1	811	29	BZ577024
42	161	8.0	867	29	BZ566980
43	160	7.9	1191	25*	BZ552701
44	159	7.9	615	29	BZ345306
45	158.5	7.9	1083	29	BZ579554

ALIGNMENTS

RESULT 1
CNS01JMC 893 bp DNA linear GSS 12-JUN-2001
LOCUS
DEFINITION Anopheles gambiae GSS T7 end of clone 13008 of NotreDanel library
from strain PEST of Anopheles gambiae (African malaria mosquito),
genomic survey sequence.
ACCESSION AL147221
VERSION AL147221.1 GI:7005367
KEYWORDS
SOURCE
ORGANISM Anopheles gambiae (African malaria mosquito)
Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;

REFERENCE	Anopheles.
AUTHORS	1 (bases 1 to 893)
TITLE	Genoscope.
JOURNAL	Direct Submission
REFERENCE	Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage
AUTHORS	BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
TITLE	- Web : www.genoscope.cns.fr)
JOURNAL	2 (bases 1 to 893)
COMMENT	Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J. Submitted (16-FEB-2000) BMJ, Institut Pasteur, 25, rue du Dr. Roux, Paris 75015, France This clone is from an A. gambiae BAC library provided by F.H. Collins and sequenced by Genoscope in collaboration with the Laboratory of Biochem. and Biol. Molec. of Insects, Institut Pasteur.
FEATURES	Location/Qualifiers
source	1..893 /organism="Anopheles gambiae" /mol_type="genomic DNA" /strain="PEST" /db_xref="taxon:7165" /clone="j3008" /clone_lib="Notredame1" /note="end : 77"
BASE COUNT	174 a 247 c 246 g 221 t 5 others
ORIGIN	
Alignment Scores:	
Pred. No.:	3,34e-51 Length: 893
Score:	585.00 Matches: 128
Percent Similarity:	64.67% Conservative: 66
Best Local Similarity:	42.67% Mismatches: 96
Query Match:	28.97% Indels: 12
DB:	29 Gaps: 2
US-09-893-756A-2 (1-412) x CNSOLJMC (1-893)	
Oy	85 LysATgLeuPhneGIngluYseTYrValArgAlaGlYcInProLeuTYrgInIleasp 104
Dd	888 AAAGCAACTTCACCGAAAGCGGTGATGTGCAGSCGGTGAGTCTGTATCAGATTGAT 829
Oy	105 SerSerThTYrGLuaIAasnLeuGIserAlAArgAlGInleuAlathralaginala 124
Dd	828 CCGCAACTATCACGCCGTATGAAGCGGAAAGCGGATCTGGCTAAGCACAGGCC 769
Oy	125 ThrLeuAlalYsAlasplAspleuAlArGYTyLyrProLeuValAlalaGuLa 144
Dd	768 GCGGCTAAAATTGGCCCC-GTGAAGCTGAACCGGTATCAAAAAAGTCCGTACCAGTAAC 710
Oy	145 ValSerArgIngluTYrAspAlaAlaValATHralYsArgSerAlaGuIaGlyVal 164
Dd	709 ATCAGTCACAGAGATTACGATACCGCCCTGGCGGAGATCCCACAGCATGAAGCGCCGTG 650
Oy	165 LysAlaAlaGlnAlaAlaleYsSerAlaGlyIleasnLeuAsnArgSerArgIlethr 184
Dd	649 GTGGCAGCAAGCGGCCGTGAAACCGGCGCATTAACCTGGCTATACCAAAGTGACC 590
Oy	185 AlaProIIeserGlyPheIleGlyGInserLYsValSerGInGlyThrLeuAsnAla 204
Dd	589 TCCTCTATCAGCGGTGATTTGGTAAATCTTCGTACGGAAGGGGCTGTGTGCAAAC 530
Oy	205 GlYAspThrThValLeuAlathrlIeaArgGInThrasnPrometTYrValAenValThr 224
Dd	528 GGTCAGANACATCGCGTGGCAACCGTCGACGACGAGYTGCATTCGATGTGACGTACAG 470
Oy	225 GInserAlaserGluValIecLYsLeuArgArgGInIleagluclyLYsLeuAla 244
Dd	469 CAGTCAGCAAGATGATTTCTCGCGCTGAAACAGAGACTGGCTAACGGCACCTGAAACAG 410
Oy	245 AlaSepGlyValIleAlaValGlyIelYsPheAspAspGlyThrValTYrProGluLYs 264
Dd	409 GAAGAAGCGGAAGCCAAAGTGAAGCTGATTACCAAGACGGTATCAAGTTCCCGCAGAA 350

OY	265	GLYVRGGLLEUENPHEALAAAPROVALVALASGLSErthrgygnliemthLeuA	284
Db	349	GGGACGCTGGAGATTCCTGACGTACCGATCGACGACCGGTTCCATCACCTTACGT	290
OY	285	AlAAlaValPProAsnaBsgInAsnIleLeuMetProGlyLeuThyValArgValLeuMet	304
Db	289	GGCATTTTCCCGAACCTGACAAAAATCTGCTGCACAGTATGTTGTTCCGACACGTCGT	230
OY	305	AspGlnValAlaValAspAsnAlaPheValAlPProGlnGlnAlaValThrArg-----	322
Db	229	GAAGAAGAAACGAATCCACCGCCCTTCGTGTTCCACAGCAGGGGTGTGACCCGTASCCAC	170
OY	323	GLYValAlaAspThrValMetIleValAsnAlaGlnGlyGlyMetGluProArgGluVal	342
Db	169	GGCGCGATGCGC-AACGCACATGGTTGTTGGGCGTCATACAAGTCGAAATCGCAATATC	111
OY	343	ThyValAlaGlnGlnGlnGlyThrAsnTrpIleValThrSerGlyLeuLysAspGlyAsp	362
Db	110	ACGCCCAACCCACGGGATTTGGGATTAATGGCTGTGTGACGGAAGGTCGAAAGATGGCGAT	51
OY	363	LysValValValAlaGluGlyIleSerIleAlaGlyIleThrGlyAlaLysValThrPro	382
Db	50	CGCGCTGATGTT-----ACTGGTTGCAAAAAATTCGTCCT	15
RESULT 2			
CONSOLJPX/C			
LOCUS	CONSOLJPX 834 bp DNA linear GSS 12-JUN-2001		
DEFINITION	Anopheles gambiae GSS T7 end of clone 14B21 of NotreDame1 library from strain PEST of Anopheles gambiae (African malaria mosquito), genomic survey sequence.		
ACCESSION	AL147350		
VERSION	AL147350.1 GI:7005496		
KEYWORDS	GSS.		
SOURCE	Anopheles gambiae (African malaria mosquito)		
ORGANISM	Anopheles gambiae Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.		
REFERENCE	1 (bases 1 to 834)		
AUTHORS	Genoscope.		
TITLE	Direct Submission		
JOURNAL	Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr		
REFERENCE	2 (bases 1 to 834)		
AUTHORS	Roch,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.		
TITLE	Direct Submission		
JOURNAL	Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr. Roux, Paris 75015, France		
COMMENT	This clone is from an A. gambiae BAC library provided by F.H. Collins and sequenced by Genoscope in collaboration with the Laboratory of Biochem. and Biol. Molec. of Insects, Institut Pasteur.		
FEATURES	Location/Qualifiers		
source	1..834		
	/organism="Anopheles gambiae"		
	/mol_type="genomic DNA"		
	/strain="PEST"		
	/db_xref="taxon:7165"		
	/clone_id="14B21"		
	/clone_lib="NotreDame1"		
	/note="end : T7"		
BASE COUNT	156 a	227 c	229 g 214 t 8 others
ORIGIN			
Alignment Scores:			
Pred. No.:	1,31e-43	Length:	834
Score:	512.50	Matches:	116
Percent Similarity:	63.57%	Conservative:	62
Best Local Similarity:	41.43%	Mismatches:	91
Query Match:	25.38%	Indels:	13
DB:	29	Gaps:	2

US-09-889-756a-2 (1-412) x CNS01JPK (1-834)

QY 106 SerThrTYrGluAlaAsnLeuGluSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThr 125
 Db 831 GCACCTTATCGAGCGCTTATGMAAGCCGGAAGCCGATCTGGCTTAAAGCAGAGCCCGC 772

QY 126 LeuAlaLysAlaAspAlaAsp-LeuAlaArgTYrLysProLeuValAlaAlaGluAlaVala 145
 Db 771 GHTAAATTTGCCCCMSTGACGCTGAACCGCTATCAAAATGCTCGGTACMAAGTACAT 712

QY 145 LSerArgGlnGluTYrAspAlaAlaValAlaThrAlaLysArgSerAlaGluAlaGlyVally 165
 Db 711 CAGTCAGAGAGATTACGATTACCGCCCTGCGGATGCCAGACAGATTACCGCCGCTGCT 652

QY 165 AlaAlaGlnAlaAlaLysSerAlaGlyLysAsnLeuAsnArgSerArgLysLeuAla 185
 Db 651 GCGAGCCAAAGCGCGCTGMAACCGCGCATTTAATCTGGCTTATACMAAAGTACCTC 592

QY 185 AProLysSerGlyPheIleGlyGlnSerLysValSerGluTYrThrLeuLeuAsnAlaG 205
 Db 591 CCTATCAGCGCGTGTATGTTAAATCTTCCTCATCGGAAGGGGCTGTGTGCAAGACGG 532

QY 205 YAspThrTYrValLeuAlaThrIleArgGlnThrAsnProMetTYrValAsnValThrG 225
 Db 531 TCAGACCACTGCGCTGCAACCGCGACACCTGATCCGATCTATGTGACGTACGCA 472

QY 225 NSerAlaSerGluValMetLysLeuArgArgGlnIleAlaGluTYrLysLeuAlaAla 245
 Db 471 GTCCAGCATATATCTTCGCGCTGMAACGAGAGCTGCTACGCGCATTCGGAACAGGA 412

QY 245 AAspGlyValIleAlaValGlyLysPheAspAspGlyTYrValTYrProGluLysG 265
 Db 411 AAACGCCAA-CCCAAAGTGAGAGTGATTAACAAGAGCGATCAAGTTCCCGCAGGAAG 353

QY 265 YArgLeuLeuPheAlaAspProValValAsnGluSerThrGlyGlnIleThrLeuAlaG 285
 Db 352 GACCTGGAATTCCTGACGTGACGACGACCAACACCGCTTCAATCACTTACCTGTC 293

QY 285 AlaValProAsnAspGlnAsnIleLeuMetProGlyLeuTYrValArgValLeuMetAs 305
 Db 292 GATTTTCCCGAACCCTGCAAAAATCTGCTGCCAGATGTTCTTCCGCGCATCTGGA 233

QY 305 GlnValAlaValAspAsnAlaPheValValProGlnGlnAlaValThrArg-----G 323
 Db 232 AGAAGGAACGAATCAACGCCCTTCTGTTCCACAGAGGCTGACCCCTACCCACGC 173

QY 323 ValAlaAspThrValMetIleValAsnAlaGlnGlyLysGluProArgGluValTh 343
 Db 172 GCGCATGCG-AGCCACATGTTGTTGGGGCTGATTAACAAGTCAAAATGCGCATATCAC 114

QY 343 ValAlaGlnGlnGlnGlyThrAsnTPPIleValThSerGlyLeuLysAspGlyAsp 363
 Db 113 CGCACCCAGCGGATGCGGCTGTAATGCTGTGACCGAAGTCTGAAAGATGGCGATCG 54

QY 363 sValValValGluGlyLysSerIleAlaGlyLysThrGlyAlaLysValThrPro 382
 Db 53 CGTATGTT-----ACTGTTTGCAAAAAGTTCCTCT 20

RESULT 3
 BH368028 642 bp DNA linear GSS 10-DEC-2001
 LOCUS AG-ND-138C13.TR ND-TAM Anopheles gambiae genomic clone AG-ND-138C13
 DEFINITION
 ACCESSION
 VERSION BH368028.1 GI:17314072
 KEYWORDS GSS.
 SOURCE Anopheles gambiae (African malaria mosquito)
 ORGANISM Anopheles gambiae
 Eukaryote; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
 Anopheles.
 1 (bases 1 to 642)

AUTHORS Hong, Y.S., Hogan, J.R., Wang, X., Sarkar, A., Sim, C., Loftus, B.J., Ren, C., Huff, E.R., Canille, J.L., Black, K., Zhang, H.-B., Gardner, M.J., and Collins, F.H.

TITLE Construction of a BAC library and generation of BAC end sequence-tagged connectors for genome sequencing of the African malaria mosquito *Anopheles gambiae*

JOURNAL Mol. Genet. Genomics 268 (6), 720-728 (2003)

MEDLINE 22542063

PUBMED 12655398

COMMENT Other GSSs: AG-ND-138C13.TF
 Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: b.loftus@igr.org
 This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.
 Seq primer: M13 Rev
 Class: BAC ends.

FEATURES
 source location/qualifiers
 1..642
 /organism="Anopheles gambiae"
 /mol_type="genomic DNA"
 /strain="PEST"
 /db_xref="taxon:7165"
 /clone="AG-ND-138C13"
 /clone_lib="ND-TAM"
 /note="Vector: pECBAC1, Site_1: HindIII"

BASE COUNT 168 a 178 c 161 g 135 t

ORIGIN

Alignment Scores:
 Pred. No.: 3,97e-33 Length: 642
 Score: 411.50 Matches: 91
 Percent Similarity: 66.15% Conservative: 36
 Best Local Similarity: 47.40% Mismatches: 60
 Query Match: 20.38% Indels: 5
 DB: 28 Gaps: 1

US-09-889-756a-2 (1-412) x BH368028 (1-642)

QY 11 AlaAlaAlaLeuAlaAlaAlaValAlaLeuValLeuSerSerCysGlyLysGlyAsp 30
 Db 75 GCGGCGATTGTGATGCTTTAGCAGCGCTTGTCAACAGAGATGAACGATTAAGAAACC 134

QY 31 AlaAlaGlnGlyLysGlnProAlaGlyArgGluAlaProAlaProValAlaGlyValAla 50
 Db 135 CAACAGAAAGCGCCCAAG-----GCACCGGCTCCGGAAGGGGTGATG 179

QY 51 ThrValHisProGlnThrValAlaLeuThrValGluLeuProGlyArgLeuGluSerLeu 70
 Db 180 ACATTGAGGCAACACCTCTCAACATCACACCGATCTTCTGCTGATCCGCTGCTAT 239

QY 71 ArgThrAlaAspValArgAlaGlnValGlyLysIleGlnLysArgLeuPheGlnGlu 90
 Db 240 CGTATGCCGAATTCGCTCTCAGGTAGCGGATTTCTGAAACGTAACCTTCGTAGAA 299

QY 91 GlySerTYrValArgAlaGlyGlnProLeuTYrGlnIleAspSerSerThrTYrGluAla 110
 Db 300 GCGAGCATATCAAGCAGAGTACTTCCCTATCAAAATGATCTCTGATCTATCAGGCA 359

QY 111 AsnLeuGluSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaLysAlaAsp 130
 Db 360 GCCTATGACAGCGCTTAAAGCGATTTGGCGAAAGCCCAAGCCAGCGCTATCGCAGCG 419

Oy 131 AlaaspleuAlaAArgTyrLeuProLeuValAlaAlaGluAlaValSerArgIngluTyr 150
 Db 420 GTGACGGTTACCGCTTACCAACCATTCGCTGGTACCAGTTACATGATGACCAATTC 479
 Oy 151 AspAlaAlaValThrAlaLysArgSerArgLysAlaGluValAlaAlaGluAla 170
 Db 480 GACACCGGAGCTCCACACCTGCAGCAGCGGAGCGGCTGTGTGCTGCCAAGCCGCT 539
 Oy 171 IleLysSerArgLysIleAsnLeuAsnArgSerArgLysThrAlaProLysSerGly 190
 Db 540 GTGAAACCGCGCGCATCAACCTGCTTACACCAAGGTGACTTACCAATCTCCGCGCT 599
 Oy 191 IleGlySerLysValSerGluGlyThrLeuLeu 202
 Db 600 ATCGCAATCTCCCTGACCAAGGCGACCTGTC 635
 RESULT 4
 LOCUS BH190105 601 bp DNA linear GSS 19-OCT-2001
 DEFINITION ATXOD17TF ATXO Arabidopsis thaliana genomic clone ATXOD17, genomic
 survey sequence.
 ACCESSION BH190105
 VERSION BH190105.1 GI:16302944
 KEYWORDS GSS.
 SOURCE Arabidopsis thaliana (chale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 Kaul, S., Town, C.D., Bowman, C.L., Van Aken, S.E., Utechtack, T.V.,
 Feldblyum, T.V. and Fraser, C.M.,
 Survey sequencing of Arabidopsis thaliana BAC T24G4
 Unpublished
 Other GSSs: ATXOD17TR
 Contact: Chris Town
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 From Wash. U contig 1142. Caution: the DNA in this BAC may be from
 some non-Arabidopsis source
 Seq primer: TP
 Classes: sheared ends.
 FEATURES
 source Location/Qualifiers
 1..601
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="ATXOD17"
 /clone_id="ATXO"
 /note="Vector: pHOS2; Site 1: BstXI; 2-3 kb sheared BAC
 DNA inserted into pHOS2 using BstXI linkers"
 BASE COUNT 98 a 206 c 175 g 122 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 7.54e-29 Length: 601
 Score: 370.50 Matches: 72
 Percent Similarity: 66.11% Conservative: 47
 Best Local Similarity: 40.00% Mismatches: 58
 Query Match: 18.35% Indels: 3
 Gaps: 2
 US-09-889-756a-2 (1-412) x BH190105 (1-601)
 Oy 193 GlnSerLysValSerGluGlyThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThr 212
 Db 600 AAGCGCAGATCAGCAGAGGTCGACTGATGAGCGCAGTATGAGACCAACCGCTCGCGGCG 541

Oy 213 IleArgInThrAsnProMetTyrValAsnValThrGlnSerAlaSerGluValMetLys 232
 Db 540 ATCAACAATCCAGCCGATATTTGACTTCACGCAAGTGCAGCTCGATCTCTGCGG 481
 Oy 233 LeuArgArgGlnIleAlaGluGlyLysLeuLeuAlaAlaAsp---GlyValIleAlaVal 251
 Db 480 CTGAAGCGGGGAGGAAGAGAGCGGAACCTTCCGAGGTGCGCCGGGAAGACGCGCAGC 421
 Oy 252 GlyIleLysPheAspPargLysThrValTyrProGluLysGlyArgLeuPheAlaAsp 271
 Db 420 AAGCTGCTGTGGAAAGAGCGGACCGAGTATTCAGACCGGCAAGATCTCTTCCGAA 361
 Oy 272 ProValValAsnGlnSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAspGln 291
 Db 360 GCGGTGAGTATGAGACGACCGGATGTATCGTCCGCGGAGTTCCCAATCCGAG 301
 Oy 292 AsnIleLeuMetProGlyLeuTyrValArgValLeuMetAspGlnValAlaAlaAspAsn 311
 Db 300 AAGATCCTGCTTCCCGCATGTTCGCCGCGTGGGTGTGCGCGGCTGAGAGAGAGAC 241
 Oy 312 AlaPheValValProGlnGlnAlaValThrArg-----GlyAlaLysAspThrValMet 329
 Db 240 GTGCTACCGGTCCACGCGCGCGGTATCCGATCCAGAGCGGTCCGCGCAGCTGATG 181
 Oy 330 IleValAsnAlaGlnGlyGlyMetGluProArgGluValThrValAlaGlnGlnGly 349
 Db 180 GTCATCAGCATTTCCACCAACGCGCAGATCCGATGATTCAGATCAACACCGGTGGG 121
 Oy 350 ThrAsnIleValIleThrSerGlyLeuLysAspGlyAspLysValIleValGlyIle 369
 Db 120 GACAAGTGGGTGTGACCTCGCGGCTCAAGGCGGAGAGAAGATCATCATGAGGCTTC 61
 RESULT 5
 LOCUS BZ558375 1096 bp DNA linear GSS 17-DEC-2002
 DEFINITION pacs1-60_968.s1 pacs1-60 Pseudomonas aeruginosa genomic clone
 pacs1-60_968, genomic survey sequence.
 ACCESSION BZ558375
 VERSION BZ558375.1 GI:27172643
 KEYWORDS GSS.
 SOURCE Pseudomonas aeruginosa
 ORGANISM Pseudomonas aeruginosa
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 Pseudomonadaceae; Pseudomonas.
 1 (bases 1 to 1096)
 REFERENCES
 AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
 Burns, J.L., Kaul, R. and Olsen, M.V.
 TITLE Whole-Genome-Sequence variation among multiple isolates of
 Pseudomonas aeruginosa library
 JOURNAL J. Bacteriol., (2002) in press
 COMMENT Contact: Chris K. Raymond
 Genome Center
 University of Washington
 Box 352145, Seattle, WA 98105-2145, USA
 Tel: 2062216954
 Fax: 2066857244
 Email: craymond@u.washington.edu
 Classes: shotgun.
 FEATURES
 source Location/Qualifiers
 1..1096
 /organism="Pseudomonas aeruginosa"
 /mol_type="genomic DNA"
 /strain="1-60"
 /db_xref="taxon:287"
 /clone="pacs1-60_968"
 /clone_id="pacs1-60"
 /note="clinical isolate 1-60 whole genomic shotgun
 library."
 BASE COUNT 181 a 385 c 351 g 179 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.35e-25 Length: 1096

Score:	543.00	Matches:	83
Percent Similarity:	55.84%	Conservative:	46
Best Local Similarity:	35.93%	Mismatches:	99
Query Match:	16.99%	Indels:	3
DB:	29	Gaps:	2

```

Db 786 GTTAGCAGATCCGAC-TACTTGAACCTTTCAGCGNNGNG-----AAA 835
Qy 223 LeuValGArgGlnIleAlaGluGlyLys-----LeuAlaAlaAspGlyValIle 249
Db 836 TTGGCGCATTCGAAMGGGATCCCGGAAGCAGTGAAGTGTTGCCAAGAAATTN 895
Qy 250 AlaValGlyIleLysPheAspAspGlyThrValTyrProGluLysGlyArgLeuLeuPhe 269
Db 896 CCTGTCGCTGTTCTCTTGGCAGACCGGAGAAAGTACCCTGTGACGGAGAGTGTGTTT 955
Qy 270 AlaAspProValValaAsnGluSerThrGlyGlnIleThrLeuValAlaAlaValProAsn 289
Db 956 TCGAACCTGAGNGTGAACCCGAGACCGAACCATATGATGATGAGCCGTTGACCAAT 1015
Qy 290 AsgGlnAsnIleLeuMetPro 296
Db 1016 ACCCATTTGAATTTGTGCCC 1036

RESULT 7
B2550552/c 775 bp DNA linear GSS 17-DEC-2002
DEFINITION pasci1-60_2713.y2 pasci1-60 Pseudomonas aeruginosa genomic clone
ACCESSION B2550552
VERSION B2550552.1 GI:27154133
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 775)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-genome-Sequence Variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL J. Bacteriol., (2002) In press
COMMENT Genome Center
Contact: Chris K. Raymond
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

FEATURES
source Location/Qualifiers
1..775
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="1-60"
/db_xref="taxon:287"
/clone="pasci1-60_2713"
/clone_lib="pasci1-60"
/notes="Clinical isolate: 1-60 whole genomic shotgun
library."
BASE COUNT 120 a 247 c 252 g 154 t 2 others
ORIGIN
Alignment Scores: 7.46e-23 Length: 775
Pred. No.: 315.00 Matches: 77
Score: 53.33% Conservative: 43
Percent Similarity: 34.22% Mismatches: 89
Best Local Similarity: 15.60% Indels: 16
Query Match: 29 Gaps: 6
DB: 29

US-09-889-756a-2 (1-412) x B2550552 (1-775)
Qy 192 GlnGlnSerLysValSerGlnGlyThrLeuLeuAsnAlaGlyAspThrThrValLeuAla 211
Db 760 GGCCGCTCCCGCTTACGAAAGGCGCCGTGTGACACCGCCGCAACCGCAATGCCCC 701
Qy 212 ThrIleArgGlnThrAsnPro---MetTyrValaAsnValThrGlnSerAlaSerGluVal 230

```

```

Db 700 ACCGTGACACGAGCTTCGACCGGATCTACGTCAACGTCACACCGCGTCCACCGCCCTG 641
Qy 231 MetLysLeuArgGlnIleAlaGluGlyLysLeu---LeuAlaAlaAspGlyValIle 249
Db 640 CTGGCGCTGCNCCTCGAAGTGGCAGCGGCCAGTTGAGCGCGCCGCGACACCGCCGCG 581
Qy 250 AlaValGlyIleLysPheAspAspGlyThrValTyrProGluLysGlyArgLeuLeuPhe 269
Db 580 AAGTCTCCCTGAAGCTGAGAGACGTAGCCAAATACCCTGTGAAGTCCGCTCGAATTTC 521
Qy 270 AlaAspProValValaAsnGluSerThrGlyGlnIleThrLeuValAlaAlaValProAsn 289
Db 520 TCCAGGTTTCCGTGACGAAAGCACCAGCTCGCTGCAATCCGCGCTGTTCCCAAC 461
Qy 290 AsgGlnAsnIleLeuMetProGlyLeuTyrValArgValLeuMetAspGlnValAlaVal 309
Db 460 CCGAACACGAGCTGCTCCGCGCATGTCTTACCGCGCAGTTGACGAGAGCGCTCAAG 401
Qy 310 AspAsnAlaPheValValProGlnGlnAlaValThrArgGlyAlaLysAsp-----Thr 327
Db 400 CAGAGGCGCATCTCTCGCTCCGACGACAGCGCTGACCCGACCTCAAGCGCAGGCTACC 341
Qy 328 ValMetIleValaAsnAlaGlnGlyMetGluProArgGluValThrValAlaGlnGln 347
Db 340 GCGCTGCTGTGAACGCGCAGACACAGATGACGCTGCGGGGATGATCAAGCGCCGAGTG 281
Qy 348 GlnGlyThrAsnThrPheValThrSerGlyLeuLysAspGlyAspLysValValGlu 367
Db 280 ATCGGCGACAGATGCTGTGTTACCGAAGCCTGAACCGCGCGACAGATATTAACGGA 221
Qy 368 GlyIleSerIleAla-----GlyIleThrGlyAlaLysLysValThrProLysGluTrp 385
Db 220 GGCTCGCAGTTCGTGCGACCGCGGTGTGAG---CTGAAGACCGTCCGCGCAAGAATGTC 164
Qy 386 AlaSerSerGluAsnGlnAlaAlaIleProGlnSerGlyValGlnThrAlaSerGluAla 405
Db 163 GCGTCCGCGCAGAGCGCGACGCGCTCCG-----GCG 131
Qy 406 LysThrAlaSerGlu 410
Db 130 AAAACGACAGCAAG 116

RESULT 8
BH190022 515 bp DNA linear GSS 19-OCT-2001
DEFINITION ATXOD32TF ATXO Arabidopsis thaliana genomic clone ATXOD32, genomic
survey sequence.
ACCESSION BH190022
VERSION BH190022.1 GI:16302757
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicot; rosids
; eurosids II; Brassicales; Brassicaceae; Arabidopsie.
REFERENCE 1 (bases 1 to 515)
AUTHORS Kaul,S., Town,C.D., Bowman,C.L., Van Aken,S.E., Utterback,T.V.,
Feldblum,T.V. and Fraser,C.M.
TITLE Survey sequencing of Arabidopsis thaliana BAC T2464
JOURNAL Unpublished
COMMENT Other GSSs: ATXOD32TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
From Wash. U contig 1142. Caution: the DNA in this BAC may be from
some non-Arabidopsis source
Seg primer: TF
Class: sheared ends.
Location/Qualifiers

```


Oy 205 YAspThrValIeuAlaTrnIleArgGlnThrAsnPro 218
 Db 130 CGATACCAAGCGCGCTGTGTGATACACCGAGGTGACGCG 91
 RESULT 10
 LOCUS BH250378 798 bp DNA linear GSS 26-NOV-2001
 DEFINITION BOCAB19TF Boga Brassica oleracea genomic clone BOCAB19, genomic survey sequence.
 ACCESSION BH250378
 VERSION BH250378.1 GI:17074104
 KEYWORDS GSS.
 SOURCE Brassica oleracea
 ORGANISM Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids II; Brassicales; Brassicaceae; Brassica.
 REFERENCE 1 (bases 1 to 798)
 Town, C.D., Van Aken, S., Uteback, T., Koo, H. and Fraser, C.M. Whole genome shotgun sequencing of Brassica oleracea
 TITLE Unpublished
 JOURNAL Other_GSSs: BOCAB19TR
 COMMENT Contact: Chris Town
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TF
 Class: sheared ends.
 FEATURES
 source Location/Qualifiers
 1..798
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /strain="T01000DH3"
 /db_xref="taxon:3712"
 /clone="BOCAB19"
 /note="Vector: pHOSt1, Site 1: BstXI, 2-3 kb sheared genomic DNA inserted into pHOSt1 using BstXI linkers"
 BASE COUNT 149 a 246 c 238 g 165 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 3,766-19 Length: 798
 Score: 280.00 Matches: 66
 Percent Similarity: 62.88% Conservative: 17
 Best Local Similarity: 50.00% Mismatches: 41
 Query Match: 13.87% Indels: 8
 DB: 28 Gaps: 1
 US-09-889-756a-2 (1-412) x BH250378 (1-798)
 Oy 16 AlaAlaValAlaLeuValIeuSerSerCysGlyGlyGlyAlaAspAlaGlnGlyGly 35
 Db 425 GCGACGCGTCCGCTTCTCTGCGCGCGCGGAA----- 460
 Oy 36 GlnProAlaGlyArgGlnAlaProAlaProValAlaGlyValValThrValHisProGln 55
 Db 461 AAGAGATCGCGTCCGCCCTCAAGCGCGGAAAGCGGCGTGCATCCGTCAGCGGAG 520
 Oy 56 ThrValAlaLeuThrValGlnLeuProGlyArgLeuGlnSerLeuArgThrAlaAspVal 75
 Db 521 CCGGCGCGGCTGCTCGAAGCTGCGGCGCGCTACGAGCGCTATCTGCGCGAGGTG 580
 Oy 76 ArgAlaGlnValGlyGlyIleGlnIleArgLeuPheGlnGlnGlySerThrValArg 95
 Db 581 CCGGACGCGGTGACGATCTGTTGGCGCGGCTTACGAGGAGGACGACCTCAAG 640
 Oy 96 AlaGlyGlnProLeuThrGlnIleAspSerSerThrTyrGlnAlaAsnLeuGlnSerAla 115

Db 641 GCCGCGACGCGCTGTACAAATCGATCCGACCCCTATATCGCGCAATTGAACGCGCG 700
 Oy 116 ArgAlaGlnLeuAlaTrnAlaGlnAlaThrLeuAlaIleAspAlaAspLeuAlaArg 135
 Db 701 AAGCGACGCTCCGGAAGCGCGGAGCGCAACCTCCGACGACGAGAACGCGCTGTGCGCGC 760
 Oy 136 TyrIleProLeuValAlaAlaGlnAlaValSerArg 147
 Db 761 TACAAAGTGTCTGTCGCGCGGACGCGCGTACGAG 796
 RESULT 11
 LOCUS BH189811/c 620 bp DNA linear GSS 19-OCT-2001
 DEFINITION ATXOC54TF ATXO Arabidopsis thaliana genomic clone ATXOC54, genomic survey sequence.
 ACCESSION BH189811
 VERSION BH189811.1 GI:16302308
 KEYWORDS GSS.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1 (bases 1 to 620)
 Kaul, S., Town, C.D., Bowman, C.L., Van Aken, S.E., Uteback, T.V., Feldblyum, T.V. and Fraser, C.M. Survey sequencing of Arabidopsis thaliana BAC T24G4
 TITLE Unpublished
 JOURNAL Other_GSSs: ATXOC54TR
 COMMENT Contact: Chris Town
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TF
 Class: sheared ends.
 FEATURES
 source Location/Qualifiers
 1..620
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="ATXOC54"
 /note="Vector: pHOSt2, Site 1: BstXI, 2-3 kb sheared BAC DNA inserted into pHOSt2 using BstXI linkers"
 BASE COUNT 108 a 191 c 199 g 122 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 3,46-18 Length: 620
 Score: 269.50 Matches: 65
 Percent Similarity: 61.29% Conservative: 30
 Best Local Similarity: 41.94% Mismatches: 47
 Query Match: 13.35% Indels: 13
 DB: 28 Gaps: 3
 US-09-889-756a-2 (1-412) x BH189811 (1-620)
 Oy 20 LeuValIeuSerSerCysGlyGlyGlyAlaAspAlaGlnGlyGlnProAlaGly 39
 Db 430 CTGCTGCTGCGCTGCGCAAGAGACGACGCGC-----GGCGCGCG----- 386
 Oy 40 ArgGlnAlaProAlaProValAlaGlyValValThrValHisProGlnThrValAlaLeu 59
 Db 385 ---CAATGCGCGCTGCGCGGCTGCTCTGCGCGGCGGACGCAAGCGTACGATC 329
 Oy 60 ThrValGlnLeuProGlyArgLeuGlnSerLeuArgThrAlaAspValArgAlaGlnVal 79
 Db 328 ACCGCGAACTTCCGCGCGGATGATCGGTGAGGTGCGGAGGTGCGCGCGCTC 269

QY 80 GlyGlyIleIleGlnLysArgLeuPheuglnGlySerTyrValArgAlaGlyGlnPro 99
 Db 268 GCGCGCATCTCTGCTGAGAAAGACTTCCAGGAAGCATGACGTGAAGCGCGGTACCTG 209
 QY 100 LeuTyrGlnIleLeuSerSerThyTyrGlnIleAsnLeuGluSerAlaArgAlaGlnLeu 119
 Db 208 CTTTCAAGATCGATCCGCTCCGCTCCAGCGGAGAAATGCGCGCGCTCTC 149
 QY 120 AlaThrAlaGlnAlaThrLeuAlaLysAlaAspAlaAspAlaArgTyrLysProLeu 139
 Db 148 GCAAGCGCGCGAGCCAACTTGACACGAGCAAGCAGCTTGATCGCTACGAGTCCCTG 89
 QY 140 ValAlaAlaGlnAlaValSerArgGlnGluTyrAspAlaAlaValThrAlaLysArgSer 159
 Db 88 GTGACTTCGAACGGCGTCCAGACAGCAGGCTTATGAC-----AAT 50
 QY 160 AlaGlnAlaGlyValLysAlaAlaGlnAlaAlaIleLysSerAla 174
 Db 49 GCGAGTCCGCGGTGAAGTGGCCGACCCGAGCTCAAGCGCG 5
 RESULT 12
 BZ557583 1263 bp DNA linear GSS 17-DEC-2002
 LOCUS pac61-60_635.s2 pac61-60 Pseudomonas aeruginosa genomic clone
 DEFINITION pac61-60_635, genomic survey sequence.
 ACCESSION BZ557583
 VERSION BZ557583.1 GI:27170544
 KEYWORDS GSS.
 SOURCE Pseudomonas aeruginosa
 ORGANISM Pseudomonas aeruginosa
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 Pseudomonadaceae; Pseudomonas.
 1 (bases 1 to 1263)
 REFERENCE
 AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
 Burns, J.D., Kaul, R. and Olsen, M.V.
 TITLE Whole-Genome-Sequence Variation among multiple isolates of
 Pseudomonas aeruginosa library
 JOURNAL J. Bacteriol., (2002) In press
 COMMENT Contact: Chris K. Raymond
 Genome Center
 University of Washington
 Box 352145, Seattle, WA 98105-2145, USA
 Tel: 2062216954
 Fax: 2066857244
 Email: craymond@u.washington.edu
 Classes: shotgun.
 FEATURES
 source Location/Qualifiers
 1..1263
 /organism="Pseudomonas aeruginosa"
 /mol_type="genomic DNA"
 /strain="1-60"
 /db_xref="taxon:287"
 /clone="pac61-60_635"
 /cdate="11b="pac61-60"
 /note="clinical isolate 1-60 whole genomic shotgun
 library."
 BASE COUNT 208 a 386 c 412 g 257 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 6,26e-17 Length: 1263
 Score: 261.50 Matches: 96
 Percent Similarity: 41.72% Conservative: 45
 Best Local Similarity: 28.40% Mismatches: 125
 Query Match: 12.95% Indels: 72
 DB: 29 Gaps: 9
 US-09-889-756a-2 (1-412) x BZ557583 (1-1263)
 QY 69 SerLeuArgThrAlaAspValArgAlaGlnValGlyIleIleGlnLysArgLeuPhe 88
 Db 37 TCGTTCACAACGTGTGATCCGACCCAGATGACGCGCATGACCCGCTGCTG 96

QY 89 GlnGlnGlySerTyrValArgAlaGlyGlnProLeuTyrGlnIleAspSerThyTyr 108
 Db 97 AGCGAAGCGCAGATGCTCGAGCGCGGAGATGCTGGCGACCATCGATGACCGCGCTC 156
 QY 109 GlnAlaAsnLeuGlnSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaLys 128
 Db 157 GTCCCGCGCTGAGCAGCGCGCAGGCTCTGCGCGAGCTACCGAGCCAGCTGATATCC 216
 QY 129 AlaAspAlaAspLeuAlaArgTyrLysProLeuValAlaAlaGlnAlaValSerArgGln 148
 Db 217 GCCAGAGAACCTCGAACCTGACCGAGCTGTATCCAGCGCTGCTCATGCGCAA 276
 QY 149 GluTyrAspAlaAlaValThrAlaLysArgSerAlaGlnAlaGlyValLysAlaAlaGln 168
 Db 277 CTGTGTGACTACGACGAGCGGCGGTGACCATGTTGGCGGCTCTGAAAGCCAAAGAT 336
 QY 169 AlaAlaIleLysSerAlaGlyIleAsnLeuAsnArgSerArgIleThrAlaProIleSer 188
 Db 337 GCCACCATCAACCGCGAGCGGCTGCGCTGTCTTACACCGGATCATCTCGCGGTATCC 396
 QY 189 GlyPheIleGlyGlnSerArgValSerGlnGlyThrLeuLeuAsnAlaGlyAspThrThr 208
 Db 397 GGCAGAGTCCGTATCCGCAACGTGATGCGCAACCTGTGCGGCGGCGACAGTCTC 456
 QY 209 ValLeuAlaThrIleArgGlnThrAspProMetTyrValAsnValThrGlnSerAlaSer 228
 Db 457 GGCCTGTTCAGCGTACCGCCAGATGACCGCATCTCTCTTCTCTCCGCAACAGAA 516
 QY 229 GluValMetLysLeuArgArgGlnIleAlaGlnGlyLysLeuLeuAlaAlaAspGlyVal 248
 Db 517 CAGTGGCCCGAGTTGCAG-----GCGCTGCTCGCGCGGCGACGCGGCG 558
 QY 249 Ile-----AlaValGlyIleLysPheAspAspGlyThr 259
 Db 559 GTGCGCGCTTACAGCCGCGACCGCGGAGCGCGCTGGC-----597
 QY 260 ValTyrProGlnLysArgLeuLeuPheAlaAspProValValAsnGlnSerThrGly 279
 Db 598 -----CGAGGCGGTTGACAGACCATGACCAACAGATTCGCAATTCCACGCT 645
 QY 280 GlnIleThrLeuArgAla-----AlaValProAsnAspGlnAsnIleLeu 294
 Db 646 ACCATCGCGGTGGCGCGCTCTCTATCGATACCGCCAGCGCGC-----TTC 693
 QY 295 MetProGlyLeuTyrValArgValLeuMetAspGlnValAlaValAspAsnAlaPheVal 314
 Db 694 TTCGCCGCGGAGTTCGCTCGCGCTTACGCTTCA-----729
 QY 315 ValProGlnGlnAlaValThrArgGlyAlaLysAspThrValMetIleValAsnAlaGln 334
 Db 730 -----CACCGGCGCTC-----741
 QY 335 GlyGlyMetGlnProArgGlnValThrValAlaGlnGlnGlnGlyThrAsnTyrIleVal 354
 Db 742 ---GGCGGGAACA-----GTTGGTCTCTGTCGAGCAAAAGCGTTGCG 786
 QY 355 ThrSerGlyLeuLysAspGlyAspLysValValGlnGlyIleSerIleAlaGlyIle 374
 Db 787 GCGCGCGCGCTGTGTGAAGGAATGTCCTTACGCGGGGTAAACGCTCCCGCACA 846
 QY 374 eThrGlyAla-----LysLysValThrProLysGln 384
 Db 847 CACGGGTCTTAAGCGCGGTTGGCCAAAGTTCCACGCGTTCCTCCAAAGAG 898
 RESULT 13
 BZ562652 937 bp DNA linear GSS 17-DEC-2002
 LOCUS pac62-164_3863.x1 pac62-164 Pseudomonas aeruginosa genomic clone
 DEFINITION pac62-164_3863, genomic survey sequence.
 ACCESSION BZ562652
 VERSION BZ562652.1 GI:27185018
 KEYWORDS GSS.

SOURCE	Pseudomonas aeruginosa
ORGANISM	Pseudomonas aeruginosa Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.
REFERENCE	1 (bases 1 to 937)
AUTHORS	Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE	Whole-genome-sequence variation among multiple isolates of Pseudomonas aeruginosa library
JOURNAL	J. Bacteriol., (2002) In press
COMMENT	Contact: Chris K. Raymond Genome Center University of Washington Box 352145, Seattle, WA 98105-2145, USA Tel: 20662216954 Fax: 2066857244 Email: ckraymond@u.washington.edu Class: shotgun.
FEATURES	Location/Qualifiers
SOURCE	1..937 /organism="Pseudomonas aeruginosa" /mol_type="genomic DNA" /strain="2-164" /db_xref="taxon:287" /clone="pacs2-164.3863" /clone_1fb="pacs2-164" /note="Clinical isolate 2-164 Whole genomic shotgun library."
BASE COUNT	155 a 305 c 327 g 147 t 3 others
ORIGIN	
Alignment Scores:	
Pid. No.:	2,56e-16 Length: 937
Score:	254.00 Matches: 91
Percent Similarity:	42.098 Conservative: 42
Best Local Similarity:	28.804 Mismatch: 132
Query Match:	12.584 Indels: 51
DB:	29 Gaps: 8
US-09-889-756A-2 (1-412) x BZ562652 (1-937)	
OY	99 ProLeuTYrGlnIleAspSerSerThrTYrGluAlaAsnLeuGlnSerAlaArgAlaGln 118 :::
Db	93 CCTGTAGCCGCATCATGATCAACCGCGCGTGTGTGCGCGCTGAGGAGCGCTCAGCGCTCC 152 :::
OY	119 LeuAlaThrAlaGlnAlaThrIleuAlaAlaAlaAspAlaAspLeuAlaArgTYrLysPro 138 :::
Db	153 AGGGGAGGAACCCAGGCGCACTGAAATCCGCGGAGGAGGAGCTGCAACCTTACCGGAGC 212 :::
OY	139 LeuValAlaAlaGluAlaValSerArgGlnGluTYrAspAlaAlaValThrAlaLysArg 158 :::
Db	213 CTGTATGCCGAGCGTGGGTGTGGCGCCAACTGCTGAGCAGGAGGAGCGAGCGGTGCAC 272 :::
OY	159 SerAlaGluAlaGlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeu 178 :::
Db	273 CAGTTCGGCGCAACCTCGAAGGCAACGATGCACCATCAACCCCGAGCGGTGGCGCTG 332 :::
OY	179 AsnArgSerArgIleThrAlaProIleSerGlyPheIleGlyGlnSerLysValSerGlu 198 :::
Db	333 TCCTTACACCCGGATCCCTGCGCGGATCCGGGAAGTCCGTATCCGCCAAGTCGATGTC 352 :::
OY	199 GlyThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAsnPro 218 :::
Db	393 GGCAACCTGTGTGGGTGGCGGACGATCTCGGCTGTTCAGCGTGAACCGAGATCGACCG 452 :::
OY	219 MetTYrValAsnValThrGlnSerAlaSerGluValMetLysLeuArgArgGlnIle--- 237 :::
Db	453 ATCTCGTGTCTTCTCCCTGCAACAGGAACAAGTGCACAGTGCAGGCGCTGCTGGC 512 :::
OY	238 -----AlaGluGlyLysLeuLeuAlaAlaAsnGlyValIleValaGlyIleLysPhe 255 :::
Db	513 GCGGAGGCGCGGTGCGCGCTACACCGCGGACCGGAGCGGCGCGCGTGGGC----- 563

OY		256	AspaspGlyThrValTyrProGluIleuArgLeuPheAlaASPProValValasn	275
Dd		564	-----GAGGGCCGGTTCTCGACCATCGACACCAATGCAC	599
OY		276	GluSerThrglyglnilethrLeuAArgAlaAlaValProAsnAepginAsnilleuMet	295
Dd		600	AGTTTCACCGGTACCATTCCGGCTGGCGGCTCTTCGATTAACCGCAGGCCGGCTTGG	659
OY		296	ProGlyLeuTyrValAlaGValLeuMechAspginValAlaValaAspnAlaPheValVal	315
Dd		660	CCGNGCAGGTGNNCGG-----	677
OY		316	ProGlnGlnAlaValThrArgGlyAlaAlasphrValMetIleValAsnAlaglNly	335
Dd		678	-----TGACCTCGCACCCGGGGGTGGCGCCGACAGTTGGTCTGTGCACAAAGCGG	731
OY		336	GlyMetGluPro-ArgGluValThrValAlaGlInglnglNlyThr-----AsnTr	352
Dd		732	GGCGCGCGGCTGAAGGGCATTTTCGTTAACCGGGTCCGCCGACCCGGTCGGGGCGGTG	791
OY		352	P-----IleValThrSer-----GlyLeuysAspGlyAspIysVa	364
Dd		792	GCCAGGCCGGGTCTTCAGACATCGACCGGCTAAGTGTGTGAAGGCCCGCCAGCGG--	849
OY		364	IValValGluIlyleSerIleAlaglyIleThrGlyAlaLysVaValThrProlyseGI	384
Dd		850	-----GGACCGGGGGGGGTGAAGGCCCATCCGGTGTGATGCCCCGGGGC	893
OY		384	uTPAlaseRserGluAsnGlnAlaAlaAlaProGlnserGlyVal	399
Dd		894	CTGGTCC-----AATATCAGAACC CGCCCAACTGTGTTA	930
RESULT 14		BZ574393	1417 bp DNA linear GSS 17-DEC-2002	
LOCUS		BZ574393	msh2_3648.xl msh Pseudomonas aeruginosa genomic clone msh2_3648,	
DEFINITION			genomic survey sequence.	
ACCESSION		BZ574393		
VERSION		BZ574393.1	GI:27209454	
KEYWORDS			GSS.	
SOURCE			Pseudomonas aeruginosa	
ORGANISM			Pseudomonas aeruginosa	
REFERENCE			Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;	
AUTHORS			Pseudomonadaceae; Pseudomonas.	
TITLE			1 (bases 1 to 1417)	
JOURNAL			Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,	
COMMENT			Burns,J.L., Kaul,R. and Olsen,M.V.	
			Whole-Genome-Sequence variation among multiple isolates of	
			Pseudomonas aeruginosa library	
			J. Bacteriol., (2002) In press	
			Contact: Chris K. Raymond	
			Genome Center	
			University of Washington	
			Box 352145, Seattle, WA 98105-2145, USA	
			Tel: 2062216854	
			Fax: 2066857244	
			Email: ckraymond@u.washington.edu	
			Class: shotgun.	
FEATURES			location/Qualifiers	
source			1..1417	
			/organism="Pseudomonas aeruginosa"	
			/mol_type="genomic DNA"	
			/strain="MSH"	
			/db_xref="taxon:287"	
			/clone="msh2_3648"	
			/clone_lib="msh"	
			/note="Environmental isolate. Whole genomic shotgun	
			library."	
BASE COUNT			242 a 399 c 376 g 397 t 3 others	
ORIGIN				
Alignment Scores:			1.19e-15 Length: 1417	
Pred. No.:				

Qy 159 eralaglu-Alaglyvallyalalagln-Alalalelysseralaglyleasne 178
 Db 328 CAGCGAATGCTGGCGTATCTGCGCGAATGCTCGCTGAATCGCGCATATCT 269
 Qy 178 uasnarserarg-1lethrallaprolleserglyphenilglylmerlyvalaser- 197
 Db 268 GCGTACACCATPAGTACCTCTCCGATTACCGATTGGTAAAGTGAACGAGAG 209
 Qy 198 --gluglyThleuleusnalaglyaspthrThvalleuAlathrillearglnThra 217
 Db 208 GAAGCGCACTTGGACAGTACGTCAGCGCATCTCGCTGCCAACCTGCAGCACTTG 149
 Qy 217 snpMetTyrValasnaValthrGlnSerAlaSerGluValMetLysLeuArg- 236
 Db 148 ATCCGATCTACGTGATGATGACCGCAGTCAGCAACGATCTCGCGCTGAACATGAA 89
 Qy 237 lleaAlaglugly-----Lysleuleu 243
 Db 88 CTGCGAATGGCACAGCTGAATCAAGATGACGCGCAATAGCCACAGTGCATCTGATC 29
 Qy 244 Alalaaspgly 247
 Db 28 ACCAGTACGCGC 17

RESULT 16
 BH190195 620 bp DNA linear GSS 19-OCT-2001
 LOCUS ATXOC46TR ATXO Arabidopsis thaliana genomic clone ATXOC46, genomic
 DEFINITION survey sequence.
 ACCESSION BH190195
 VERSION BH190195.1 GI:16303143
 KEYWORDS GSS.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1 (bases 1 to 620)
 Kaul, S., Town, C.D., Bowman, C.L., Van Aken, S.E., Uteback, T.V.,
 Feldblyum, T.V. and Fraser, C.M.
 Survey sequencing of Arabidopsis thaliana BAC T24G4
 TITLE Unpublished
 JOURNAL Contact: Chris Town
 COMMENT TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 From Wash. U contig 1142. Caution: the DNA in this BAC may be from
 some non-Arabidopsis source
 Seg primer: TR
 Class: sheared ends.

FEATURES
 source Location/Qualifiers
 1..620
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="ATXOC46"
 /note="Vector: PHOS2; site_1: BstXI; 2-3 kb sheared BAC
 DNA inserted into PHOS2 using BstXI linkers"

BASE COUNT 113 a 201 c 184 g 122 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1.01e-15 Length: 620
 Score: 246.00 Matches: 52
 Percent Similarity: 61.07% Conservative: 28
 Best Local Similarity: 39.69% Mismatches: 47
 Query Match: 12.18% Indels: 4
 DB: 28 Gaps: 3

US-09-889-756a-2 (1-412) x BH190195 (1-620)
 Qy 265 GlyargLeuLeuPheAlaAspProValValasnluserThrglyGlnleThreLeuArg 284
 Db 4 GCGAAGATCTCTCTTCCGAAAGCGGTGTGATAGACGACCGGAGTACCGTGGC 63
 Qy 285 AlalaValProAsnaAspGlnAsnilleuLeuMetProGlyLeuTyrValArgValleuMet 304
 Db 64 GCGAGTTTCCCATCCCGAAGATCTCTCTCCGCGCATGTTCGCGCGCGCGGTG 123
 Qy 305 AspGlnValAlaValAspaAlaPheValValProGlnGlnAlaValThraArg----- 322
 Db 124 GTGAGAGCGGTGAAGAGAGCGTGCACCGTCCACAGCGCGCGTGAACCGCATGCA 183
 Qy 323 GlyAlaLysAspThrValMetIleValAsnAlaGlnlyGlyMetGluProArgGluVal 342
 Db 184 GCGGTGCGCGGACGCGTATGTATGACAGATTCACACCGCGCAGATCCGATGATC 243
 Qy 343 ThrValAlaGlnGlnGlnlyThrAsnTrpIleValThrSerGlyLeuLysAspGlyAsp 362
 Db 244 CAGACTGACAAACGCGTGGGCGACAGTGGGTGTGACCTCGGCTCAAGCGCGAGAG 303
 Qy 363 LysValValValGlyGlyIleSerIleAlaGlyIleThrGlyAlaLysValThrPro 382
 Db 304 AAGTCAATCATGAGAGGCTC--CTGAAGCTGTCCTCCGAGAGG--CCGTGTCTCG 357
 Qy 383 LysGluTrpAlaSerSerGluAsnGlnAlaAla 393
 Db 358 GAGCCCTTCGTCGCGAAGCGGACCGCTCG 390

RESULT 17
 BH189826 596 bp DNA linear GSS 19-OCT-2001
 LOCUS ATXOC46TR ATXO Arabidopsis thaliana genomic clone ATXOC46, genomic
 DEFINITION survey sequence.
 ACCESSION BH189826
 VERSION BH189826.1 GI:16302338
 KEYWORDS GSS.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1 (bases 1 to 596)
 Kaul, S., Town, C.D., Bowman, C.L., Van Aken, S.E., Uteback, T.V.,
 Feldblyum, T.V. and Fraser, C.M.
 Survey sequencing of Arabidopsis thaliana BAC T24G4
 TITLE Unpublished
 JOURNAL Other GSSs: ATXOC46TF
 COMMENT Contact: Chris Town
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 From Wash. U contig 1142. Caution: the DNA in this BAC may be from
 some non-Arabidopsis source
 Seg primer: TR
 Class: sheared ends.

FEATURES
 source Location/Qualifiers
 1..596
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="ATXOC46"
 /note="Vector: PHOS2; site_1: BstXI; 2-3 kb sheared BAC
 DNA inserted into PHOS2 using BstXI linkers"

BASE COUNT 104 a 183 c 195 g 114 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1.01e-15 Length: 596
 Score: 246.00 Matches: 52
 Percent Similarity: 61.07% Conservative: 28
 Best Local Similarity: 39.69% Mismatches: 47
 Query Match: 12.18% Indels: 4
 DB: 28 Gaps: 3

Pred. No.:	15	length:	59
Score:	244.00	Matches:	61
Percent Similarity:	58.94%	Conservative:	28
Best Local Similarity:	40.40%	Mismatches:	48
Query Match:	12.09%	Indels:	14
DB:	28	Gaps:	3
.			

US-09-889-756A-2 (1-412) x BH189826 (1-596)

```

OY      20  LeuValIeuSerCysGlyLysGlyAlaPalaalagInGLyGLNProIalagLy 39
          |||   |||   ::|||   |||   |||   |||   |||   |||
Db      424 CTGTCTGTCCCGCTTGCCAAAAGAACAAAGGGCC-----GGCCCCCG-- 380

OY      40  ArgGluAlaProIalProValIGLYalValIalThrHisProGlnThrAlaLeu 59
          :|||   |||   |||   |||   |||   |||   |||   |||
Db      379 ----CAAATGCCCGCTGC CGGGGAlaACTTTCCTCCCGCGCGTACCAGAAACGTATAAGCATTC 323

```

Db 322 ACCCGCGAAGCTCCCGCGGATCGATCCGGTGGCGAGGTGCGCGCGCTC 2633

Db

262 GGGGCATCTCTGTGAGAACCTTTCAGGAAGGCATGACGTGAAGGGGGTGCACGTG 203

Db 202 CTTTCAAGTCGATCCGGGCTCCGCTCGAGCGCGCGCAAGAGAGATGCGCGCGGCTCTC 143

Db 142 GCACGGGCGGAGGCAACTTGACACGAGCAAGACCGAGCTTGATCGCTACGAGTCCCTG 83

[illegible]

Db 46 ATCGAAGAAACGGTGTAAAGTGGCCGACGCGAGC 14
RESUIT 18

RESULT 18					
B2577818					
LOCUS	BZ577818	1108 bp	DNA	linear	GSS 17-DEC-2002
DEFINITION	msl2_5586.y2 msh Pseudomonas aeruginosa genomic clone msl2_5586, genomic survey sequence.				

ACCESSION	BZ577818
VERSION	BZ577818.1
KEYWORDS	GI:27212879 GSS.

ORGANISM
Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales
Pseudomonadaceae; Pseudomonas.
1 (Accession 145-3100)

TITLE Whole-genome-Sequence variation among multiple isolates of *Pseudomonas aeruginosa* library

JOURNAL J. Bacteriol. 165 (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel.: 2062216594
Fax: 2066857244
Email: ckraymond@u.washington.edu
Class: Shotgun.

```
FEATURES
    source
        Location/Qualifiers
            1..1108
                /organism="Pseudomonas aeruginosa"
                /mol_type="genomic DNA"
                /strain="MSH"
                /db_xref="taxon:287"
```

BASE COUNT	ORIGIN
242 a	352 c
306 g	192 t
16 others	

Alignment Scores

Score:	241.50	Matches:	80
Percent Similarity:	56.63%	Conservative:	31
Best Local Similarity:	40.82%	Mismatches:	79
Query Match:	11.96%	Indels:	11
DB:	29	Gaps:	3

Db

Cy

290 CTGGCGGCCCTGTCCGCCCATTCTCCTGGCTGC-----GAAGAACACACGGAC 34

Db 341 ACCGGGAAGACTGCG-----GAGGCCCGCCGAG--GTCCGCGTGATGTCGCCAGG 39

Ddb

392 CCGGCGCCCTATCGGCATCACCAAGCGAGCTGCCCGGACGCTGGAAAGCTACCGCCAGGCT 45

Db 452 GAAGTGGGGGGCGGCTCGCCCGCATGTCAACCGTGGCTGTACGAGGAGGCCAGAC- 51

Db 511 GTCCGCGCCGACCGTCTTCCAGATCGACCCCTGGCC -CTTGAGGCGGNCCTGGAC 56

Db 570 ATGAGCGCGGCGCCCTGGCCCGGCGAGGCGAGNCAGCANNCGCGCGGNCGACAAAGCTC 62

[illegible]

Db 690 CAAACCGACGCGCGCAAGCCCTGNNGCANATTGCCCTTGNNCAGAGGCGAA-CTTGACAA 74

Db 749 GCCGCCCTGGG-CTTGGCTACGCCACAGGTAAAGGGGCCCATTCAGNCGCCGCGGTNT- 80

Db 807 TCCTTGTCAACCAAGGCCCTTGGCCGAAAGAATTCCCGAACGTT 854

Accession	B5580098	1259 bp	DNA	linear	GSS	17-DEC-2002
Definition	msh2_336.y2 msh Pseudomonas aeruginosa genomic clone msh2_336, genomic survey sequence.					
Accession	B5580098					
Version	B5580098.1	GI:27215159				
Keywords	GSS.					
Source	Pseudomonas aeruginosa					
Organism	Pseudomonas aeruginosa					

REFERENCE
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.
1 (bases 1 to 1259)
Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M., Burns, J.L., Kul, R. and Olsen, M.V.
TITLE
Whole-genome sequence variation among multiple isolates of

Db	229	CAAAACCCGTGTTAAATTAACGAAAAACCCCTTACCGGAGGCTTTAAACAATGCAAAAGG	286
Qy	118	GlnLeuAlaThrAlaGlnAlaThrLeuAlaValAlaAspAlaAspLeuAlaArgTyrLys	137
Db	289	AGTTTACATGCGACCGGAACCCCTTAAACGGCTCAGTTAGAGGTATGAATTAACA	348
Qy	138	ProLeuValAlaAlaGlnAlaValSerArgGlnGluTyrAspAlaAlaValThrAlaLys	157
Db	349	CCACTGTTCCAGAAATTAAGTTGTTCTGATTCAGTTTAAACCGCAAAAACAGCTTAT	408
Qy	158	ArgSerAlaGlnAlaGlyValLysAlaAlaGlnAlaAlaLeuSerZalGlyIleAsn	177
Db	409	AAATTTGGCAGGCAAAATCGCAACAGCTTAAGCAGTGTTGTCGACACAGATTAC	468
RESULT 21			
LOCUS	B2554963/c	923 bp	DNA linear GSS 17-DEC-2000
DEFINITION	pacsl-60_4947.x1 pacsl-60 Pseudomonas aeruginosa genomic clone		
ACCESSION	B2554963		
VERSION	B2554963.1		
KEYWORDS	GSS.		
SOURCE	Pseudomonas aeruginosa		
ORGANISM	Pseudomonas aeruginosa		
REFERENCE	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.		
AUTHORS	1 (bases 1 to 923) Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V.		
TITLE	Whole-Genome-Sequence variation among multiple isolates of Pseudomonas aeruginosa library		
JOURNAL	J. Bacteriol., (2002) In press		
COMMENT	Contact: Chris K. Raymond Genome Center University of Washington Box 351145, Seattle, WA 98105-2145, USA Tel: 2062316954 Fax: 2066857244 Email: ckraymond@u.washington.edu Class: shotgun.		
FEATURES	Location/Qualifiers		
source	1..923 /organism="Pseudomonas aeruginosa" /mol_type="genomic DNA" /strain="1-60" /db_xref="taxon:287" /clone="pacsl-60_4947" /clone_lib="pacsl-60" /note="Clinical isolate 1-60 whole genomic shotgun library."		
BASE COUNT	190 a 264 c 256 g 212 t	1 others	
ORIGIN			
Alignment Scores:			
Pred. No.:	1.08e-12	Length:	923
Score:	219.50	Matches:	66
Percent Similarity:	50.98%	Conservative:	38
Best Local Similarity:	32.35%	Mismatches:	89
Query Match:	10.87%	Indels:	11
Db:	29	Gaps:	5
US-09-889-756A-2 (1-412) x B2554963 (1-923)			
Qy	19	AlaLeuValLeuSerSerCyGlyLysGlyLysAspAlaAlaGlnGlyGlnProAla	38
Db	695	GCCTTACACTTCTGCTTC-----GGGATTTGGCCGCCCGGGGGCGCGCAGC	648
Qy	39	GlyArgGlnAlaProAlaProVal---ValGlyValAlaThrValHisProGlnThrVal	57
Db	647	GGC-----CCCGCAAGCGTTCCGTTCCGAGGAGTGATTCGTCG-----GTA	603
Qy	58	AlaLeuThrValGlnLeuProGlyArgGlnLeuSerLeuAlaGlnThrAlaAspValArgAla	77

FEATURES	source
Db	602 AACCCCTACGCCAAGTTACACCGCGCTCGTGAACCGCTGTCCAGACAGTAAACTGCNCGC 543
Qy	78 GlnValGlyIleIleIleGlnIlySargLeupheGingIugIySerTyrValAlargalagIy 97
Db	542 CGCGTCGCGCGGTACATTATTCAGGACGTACCGGTGCGCGAAGCGCGCTGTAGGAAGGCG 483
Qy	98 GlnProLeuTyrGlnIleAspSerSerThrTyrGluAlaAsnLeuGlnSerAlaValgala 117
Db	482 CAGCAACGTGTCCTCATCGACCCGCGCGTGTTCAGAGCGCGCAGAGATCCCGCAGGGCA 423
Qy	118 GlnLeuAlaThrAlaGlnAlaThrLeuAlaIlySAlaAspAlaAspLeuAlaIlyTyrTys 137
Db	422 CGCCTCGCGCAGAGCCGAGGCCACCGCGCTTCGCGCGCCGACCCGAACAGACGAGCCGCG 363
Qy	138 ProLeuValAlaAlaGlnAlaValSerArgGlnGlnTyrAspAlaAlaValAlaThrAlaTys 157
Db	362 GTGCTGTATGCGCGGAAGTGTGTCGCCCGGAGCGGCTTCGACAGCGCCATTCGCTCGCT 303
Qy	158 ArgSerAlaGlnAlaGlnValIlyLeAlaAlaGlnAlaAlaIleIlySerAlaIlyleAn 177
Db	302 AATGCCACCAAGGACACAGTGCATGCTTCACGGAACGCCCTTGACGCGCGCAATGTCT 243
Qy	178 LeuAsnArgSerArgIleThrAlaProIleIleSerGlyPheIleGlnSerIlyValSer 197
Db	242 ATCGCGCTTCGCTCGGTATACACACCGATCGCGTGGCTGTGGCATATGCAGGTCTT 183
Qy	198 GluGlyThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAsn 217
Db	182 GAGGGCTACTACTATGCACCAATGCG---GTGCGCGCTCTTACACAGATTGTTCGGTGTAT 126
Qy	218 PromethylVal 221
Db	125 CCACTGTACGTG 114
RESULT 22	
LOCUS	BH189931
DEFINITION	BH189931 597 bp DNA linear GSS 19-OCT-2001
ACCESSION	ATXOD60TR ATXO Arabidopsis thaliana genomic clone ATXOD60, genomic
VERSION	BH189931
KEYWORDS	BH189931.1 GI:16302558
ORGANISM	GSS.
REFERENCE	Arabidopsis thaliana (thale cress)
AUTHORS	Arabidopsis thaliana
TITLE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
JOURNAL	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
COMMENT	1 (bases 1 to 597) Bowman, C.L., Van Aken, S.E., Uterback, T.V., Kaul, S., Town, C.D., Feldblyum, T.V. and Fraser, C.M. Survey sequencing of Arabidopsis thaliana BAC T24G4 Unpublished
	Other GSSs: ATXOD60TF
	Contact: Chris Town
	7912 Medical Center Drive, Rockville, MD 20850, USA.
	Tel: 301-838-3523
	Fax: 301-838-0208
	Email: cdtown@ctigr.org
	From Wash. U contig 1142. Caution: the DNA in this BAC may be from
	some non-Arabidopsis source
	Seq primer: TR
	Class: sheared ends.
	Location/Qualifiers
	1..597
	/organism="Arabidopsis thaliana"
	/mol_type="genomic DNA"
	/strain="Columbia"
	/db_xref="taxon:3702"
	/clone="ATXOD60"
	/clone_lib="ATXO"
	/note="Vector: pHO52, Site 1: BstXI, 2-3 kb sheared BAC
	DNA inserted into pHO52 using BstXI linkers"

BASE COUNT 106 a 192 c 183 g 116 t
ORIGIN

Alignment Scores:
Pred. No.: 1.39e-12 Length: 597
Score: 216.00 Matches: 48
Percent Similarity: 58.54% Conservative: 24
Best Local Similarity: 39.02% Mismatches: 47
Query Match: 10.70% Indels: 4
DB: 28 Gaps: 3

US-09-889-756a-2 (1-412) x BH189931 (1-597)

Qy 273 ValValaengluSerThrGlyGlnIleThrLeuArgAlaValProAspAspGlnAsn 292
Db 3 GTGGGAGTGAAGACACCGGATGTGTCACGTCGCCCGGAGCTTCCATCCGAGAG 62
Qy 293 IleLeuMetProGlyLeuTyTValArgValLeuMetAspGlnValAlaValAspAsnAla 312
Db 63 ATCCGTGCTCCCGGATGTGTCGCCGCGGTGGTGAGCGCGAGGAGGAGAGCGTG 122
Qy 313 PheValValProGlnGlnAlaValThrArg-----GlyAlaValAspThrValMetIle 330
Db 123 GTCAACCGGACACGCGCGGTGACCCGATGCAAGCGGTCGCCGAGCGGTGGGTC 182
Qy 331 ValAsnAlaGlnGlyGlyMetGluProArgGluValThrValAlaGlnGlnGlnIleThr 350
Db 183 ATCGACGATTCACACCGGAGATCCGATGATCCAGATGACGACGAGCGGTCGAGC 242
Qy 351 AsnTrpIleValIleThrSerGlyLeuValAspGlyAspValValAlaGluGlyIleSer 370
Db 243 AAGTGGGTGTGACCTCCGCGCTCAAGCGCGGAGGAGAAAGTCATCATGAAAGGCTTC-- 299
Qy 371 IleAlaGlyIleThrGlyAlaValAspValThrProLysGluTrpAlaSerSerGluAsn 390
Db 300 CTGAAGGCTGCTCCCGAGAGC--CCGGTGTTCGAGACCTTCGTCGCAAGACCGAG 356
Qy 391 GlnAlaAla 393
Db 357 ACCGCTGCG 365

RESULT 23
BZ561644/c 747 bp DNA linear GSS 17-DEC-2002
LOCUS pacs2-164_3364.x1 pacs2-164 Pseudomonas aeruginosa genomic clone
DEFINITION pacs2-164_3364, genomic survey sequence.
ACCESSION BZ561644
VERSION BZ561644.1 GI:27182075
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 747)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence Variation among multiple isolates of
Pseudomonas aeruginosa library
J. Bacteriol., (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

FEATURES
source 1..747
location/Qualifiers
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"

BASE COUNT 112 a 238 c 251 g 144 t 2 others
ORIGIN

Alignment Scores:
Pred. No.: 1.89e-12 Length: 747
Score: 216.00 Matches: 56
Percent Similarity: 53.09% Conservative: 30
Best Local Similarity: 34.57% Mismatches: 74
Query Match: 10.70% Indels: 2
DB: 29 Gaps: 2

US-09-889-756a-2 (1-412) x BZ561644 (1-747)

Qy 183 IleThrAlaProIleSerGlyPheIleGlyIleSerValSerGluGlyThrLeuLeu 202
Db 582 GTTATCGCCCGATTAACGACCCCGCAGCGGTGTCACAAAGCGCGCTGTC 523
Qy 203 AsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAsnPrometTyTValAsn 222
Db 522 GCGTAGACTGCGGACACCTCTACCCGCTCAGACAGATGATCCGATCTAGTGAAC 463
Qy 223 ValThrGlnSerAlaSerGluValMetLysLeuArgGlnIleAlaGluGlyLysLeu 242
Db 462 TTCGCCAGCGCGCGGAGAGTCCGCCATGACGAGCGGCGATCCGGAAGCGCAGTG 403
Qy 243 Leu--AlaAlaAspGlyValIleAlaValGlyIleLysPheAspAspGlyThrValTy 261
Db 402 AAGGTGTGCGGACCAAGACATCTCGTGGCTGCTGCGCGGACGACGAGATAC 343
Qy 262 ProGluLysGlyValGluLeuPheAlaAspProValValaengluSerThrGlyGlnIle 281
Db 342 CGGCTGGCGCGGAGCTGCTGTCTCCGACCTGCGCGGTCGCCGACCGGACGACACATC 283
Qy 282 ThrLeuArgAlaAlaValProAspAspGlnAsnIleLeuMetProGlyLeuTyTValArg 301
Db 282 GCCATGCGTCCCTGTCCTCCGATCCGATCCGATCCGATCCGATCCGATCCGATCCG 223
Qy 302 ValLeuMetAspGlnValAlaValAspAsnAlaPheValValProGlnGlnAlaValThr 321
Db 222 GTGCGCTGCGGACGCGGAGTGAACCGACGATCAACCGTCCCGGAGCGGCTGATC 163
Qy 322 ArgGlyAlaLysAspThrValMet--IleValAsnAlaGlnGlyGlyMetGluProArg 340
Db 162 CGTACCGCCGACGTCGCGGTGTCAAGGTGTCCGACACAGGAGTGAAGAAAGCCCTG 103
Qy 341 GluVal 342
Db 102 GAGGTC 97

RESULT 24
BZ336018 568 bp DNA linear GSS 06-NOV-2002
LOCUS hz28c04.g1 MGS-Sbicolorf (JM107 adapted methyl filtered) Sorghum
DEFINITION bicolor genomic clone hz28c04 5', genomic survey sequence.
ACCESSION BZ336018
VERSION BZ336018.1 GI:24729117
KEYWORDS GSS.
SOURCE Sorghum bicolor (sorghum)
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 568)
AUTHORS Rabinowicz,P.D., O'Shaughnessy,A.L., Ballia,V., Dedhia,N.,
Katzenburger,F., King,L., Miller,B., Miller,S., Nascimento,L.,
Zutavern,T., Palmer,L., McCombie,W.R. and Marienissen,R.A.
TITLE Genomic shotgun sequences from Sorghum bicolor (methyl-filtered)
JOURNAL Unpublished
COMMENT Contact: W. Richard McCombie

Db 381 GAGCCGCCACCCAGCCGCCAGGAGCCAGGCGGCTGCCGCGACCCAGCGGCA 322
 Oy 171 IleuyserrAlaGlyIleAsnleuAsnArgserArgIleThrAlaProIleSergIlyphe 190
 Db 321 CTGAGCGGCGGCGGCTTAAGCTTACCTTACCGGATACCGCGCCCATGACGCTGCC 262
 Oy 191 IlegIyInserLysValSergIlyIleuAsnAlaGlyAspThrThrValIleu 210
 Db 261 GTCAAGCCGCGCGAGGTACCGCGCAACCTGTCACCTCGGGAG--ACCTGCTC 205
 Oy 211 AlaThrIleArgInThrAsnPromethyrrValaAsnValThrInSerrAlaSergIlyVal 230
 Db 204 ACCACCTGCTGTACAGCCGACAGGTCTACGCTTA-CTTCAGCGCGACGAGCGGTGT 146
 Oy 231 MetLysLeuArgArgGlnIleAlaGlnIlyIlySleuLeuAlaAlaAspGlyValIleAla 250
 Db 145 CCTCAAGTACGTGAGCTGCGCGCCGACGCGCGGTGCGACACGCGCAGAGCCGCGT 86
 Oy 251 ValGlyIleLysPheAspAspGlyThrValTyProGlnIySgIlyArgLeuIleuAla 270
 Db 85 GTGGGG----- 80
 Oy 271 AsproValValaAsnGlu 276
 Db 79 GATCCTACTAGTTCTAGAG 62

RESULT 26

B2551287/c 809 bp DNA linear GSS 17-DEC-2002
 LOCUS pasc1-60_3076.y2 pasc1-60 Pseudomonas aeruginosa genomic clone
 DEFINITION pasc1-60_3076, genomic survey sequence.
 ACCESSION B2551287
 VERSION B2551287.1 GI:27154868
 KEYWORDS GSS

SOURCE

ORGANISM Pseudomonas aeruginosa
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 Pseudomonadaceae; Pseudomonas.

REFERENCE

1 (bases 1 to 809)
 Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
 Burns,J.L., Kaul,R. and Olsen,M.V.
 Whole-Genome-Sequence Variation among multiple isolates of
 Pseudomonas aeruginosa library

JOURNAL

J. Bacteriol., (2002) In press
 Contact: Chris K. Raymond
 Genome Center
 University of Washington
 Box 352145, Seattle, WA 98105-2145, USA
 Tel: 2062216954
 Fax: 2066857244
 Email: craymond@u.washington.edu
 Class: shotgun.

COMMENT

FEATURES

source

1. 809
 Location/Qualifiers
 /organism="Pseudomonas aeruginosa"
 /mol_type="genomic DNA"
 /strain="1-60"
 /db_xref="taxon:287"
 /clone="pasc1-60_3076"
 /clone_11b="pasc1-60"
 /note="Clinical isolate 1-60 whole genomic shotgun library."
 BASE COUNT 138 a 241 c 261 g 169 t

ORIGIN

Alignment Scores:

Pred. No.: 4,37e-10 Length: 809
 Score: 194.00 Matches: 60
 Percent Similarity: 49.73% Conservative: 31
 Best Local Similarity: 32.79% Mismatches: 58
 Query Match: 9.61% Indels: 34
 29 Gaps: 7

US-09-889-756a-2 (1-412) x B2551287 (1-809)

Oy 42 AlaProAlaProValaIleGlyValIleValThrValHisProGlnInThrValAlaLeuThrVal 61
 Db 564 GCACCAAAACCCGCGATCAGGTG-----ACCGCAGCTTGGCCGA 523
 Oy 62 GlnLeuPro-----GlyArgLeuGlnSerrLeuArgThrAla 73
 Db 522 AAGGTCCTTCGAGACAGCCGCTCCAGCATCGGCGAGCCCTCAAGCATTCACAGGCGTG 463
 Oy 74 AspValArgAlaGlnValaIleGlyIleIleGlnIlySargLeuPheGlnGlnIlySerrTy 93
 Db 462 ACCCTACCGCGGAGGTCTCCGACGCTACGACGATGCTTCTTCGCGGACCCAG 403
 Oy 94 ValArgAlaGlnGlnProLeuTyrgInIleAspSerrThrTyrgIlyAlaAsnLeuGlu 113
 Db 402 GTGAAGCTGAGACCAACCGCTATCATAGTTGGAAGACAGCTGAGAAAGCACCCTGCGC 343
 Oy 114 SerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaValaAspAlaAspLeu 133
 Db 342 -----ACTGCCAGGCGCATCTCGGCTCGGCGAGGCGGAGTAC 304
 Oy 134 AlaArgTyrrLysProLeuValaAlaAlaGlnAlaValaSerArgInIlyTyrrAsp----- 151
 Db 303 CAGCGCGCGCGCGAAGCTGATCGGACGAGCAAGCCATCTCGAAMCGAATTGATGCTCTC 244
 Oy 152 AlaAlaValThrAlaLysArgSerAlaGlnAlaGlyValIlySargAlaAlaAlaIle 171
 Db 243 GCCCGGAGTGGGCGGACAGCCAGCCGCTC-----GCCGAGCTG 202
 Oy 172 LysSerAlaGlyIleAsnLeuAsnArgserArgIleThrAlaProIleSergIlyPheIle 191
 Db 201 AAGCGCGG-----CTGGCAGAGAAGCGGCTGCTCGGCGCTTCGCGGAGACATC 151
 Oy 192 GlyInserLysValSergIlyIleuLeuAsnAlaGlyAspThrThrValLeuAla 211
 Db 150 GGCACTCCGCGCGAGTGGAGCTGCGGACTACTCTCGCCCGG-----ACGCGATGCGC 97
 Oy 212 ThrIleArg 214
 Db 96 ACCATCAG 88

RESULT 27

CA554488 595 bp mRNA linear EST 19-NOV-2002
 LOCUS C0886C11-5N NIA Mouse Blastocyst cDNA library (Long) Mus musculus
 DEFINITION CDNA clone NIA:C0886C11 IMAGE:30033250 5', mRNA sequence.

ACCESSION

CA554488
 CA554488.1 GI:25098787
 EST.

KEYWORDS

SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.
 1 (bases 1 to 595)
 Piao,Y., Kargul,G.J., Dudekula,D.B., Qian,Y., Tanaka,T., Luo,A. and
 Ko,M.S.H.

TITLE

Systematic Analyses of NIA Mouse Blastocyst cDNA Library (Long)
 Unpublished
 Other ESTs: C0886C11-3
 Contact: Dawood B. Dudekula
 Laboratory of Genetics
 National Institute on Aging/National Institutes of Health
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
 Email: cdna@igun.970.nia.nih.gov
 Plate: C0886 row: C Column: 11
 Seq primer: M13 Reverse
 High quality sequence stop: 595
 POLYA=No.

COMMENT

FEATURES

source
 1. 595
 Location/Qualifiers
 /organism="Mus musculus"
 /mol_type="mRNA"


```

/strain="CS7BL/6J"
/db_xref="nlastr:C086C11-5N"
/db_xref="taxon:10090"
/clone="NIA:C086C11 IMAGE:30033250"
/issue_type="Blastocyst"
/dev_stage="3.5-dpc"
/lab_host="DH10B"
/clone_lib="NIA Mouse Blastocyst cDNA Library (Long)"
/notes="Vector: pSPORT1 (Invitrogen); Site 1: SalI; Site 2:
NotI. Mouse cDNA project by the Laboratory of Genetics,
National Institute on Aging (NIA), Intramural Research
Program, NIH (http://igsn.grc.nia.nih.gov/cDNA). This is
a long-transcript enriched cDNA library (Ref. Genome Res.
11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were
extracted from a pool of 20 Blastocysts. Double-stranded
cDNAs were synthesized with an oligo(dT) primer
[Invitrogen:
5'-pGACTGATCTTATCCGAGCGCCGCCCTTTTCTTTT-3'] from
0.2 ug of total RNA, treated with T4 DNA polymerase, and
purified by ethanol-precipitation. The cDNAs were ligated
to lone-linker lp-Sal4, purified by phenol/chloroform, and
separated from free linkers by Centricon 100. Then, the
cDNAs were amplified by long-range high fidelity PCR using
Ex Taq polymerase (Takara) with a primer Sal4-S. The
products were purified by phenol/chloroform and Centricon
100. The cDNAs were digested with SalI and NotI enzymes
and cloned into SalI/NotI site of pSPORT1 plasmid vector.
The DH10B E. coli host was transformed with the ligation
mixture by the standard chemical method. The average
insert size is about 2.2 kb. The library was constructed
by Yulan Piao (NIA)."

```

```

BASE COUNT      199 a      116 c      124 g      156 t
ORIGIN
Alignment Scores:
Pred. No.:      8.54e-10      Length:      595
Score:          189.50      Matches:      59
Percent Similarity: 49.13%      Conservative: 54
Best Local Similarity: 25.65%      Mismatches: 82
Query Match:    9.39%      Indels:      35
DB:             14      Gaps:      8

US-09-889-756a-2 (1-412) x CA554488 (1-595)
QY      78  GlnValGlyGlyIleGlnLysArgLeuPheGlnGlySerTyrValArgAlaGly  97
Db      3  GAAGTAAGTCGATTTTCTGTCGAATCTCAACAGAGGAGAAAGTATCTCAAGGA  62
QY      98  GlnProLeuTyrGlnIleAspSerSerThrTyrGlnLysLeuGlnSerAlaArgAla  117
Db      63  ACGGTGCTTGCAGAGATCAATGATGATCTACACGCAATATGAGAAATTCGGTGT  122
QY      118  GlnLeuAlaThrAlaGlnAlaThrLeuAlaLysAlaAspLeuAlaArgTyrLys  137
Db      123  CAATTGGATCTGCACAA-----AAAATGAAAGCTCGACTGAAAA  161
QY      138  ProLeuValAlaAlaGlnAlaValSerArgGlnGlyTyrAspAlaAlaThrAlaLys  157
Db      162  AAATGCTTATATCAATGATATCAAGCTGATATATATGCGCACTCAATCAAGTG  221
QY      158  ArgSerAlaGlnAlaGlyValLysAlaAlaGlnAlaAlaLysSerAlaGlyLeuAsn  177
Db      222  AATGCTTAAAGAGATCTGATATACCAAGCCAGATC-----  263
QY      178  LeuAsnArgSerArgIleThrAlaProIleSerGlyPheIleGlyGlnSerLysValSer  197
Db      264  ---GATGAAAACAGGTTAAAGCAGCTTTCACCGGTATATTAGGTTTGAGAAATGATCAGC  320
QY      198  GlnGlyThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAsn  217
Db      321  CCGGTCCTTATGTAAACA-----CCAGCTACCAATTTCTTGCTGATTAACAACCAAGAT  374
QY      218  ProMetTyrValAsnValThr-----GlnSerAlaSerGlnValMetLysLeuArgArg  235

```

```

Db      375  AAAGTAAGATCGATTTTACTGTTCCGAATGATATCGACGTT-----  419
QY      236  GlnIleAlaGlnGlyLysLeuLeuAlaAlaAspGlyValIleAlaValGlyIleLysPhe  255
Db      420  ---ATTGTTAAAGGAAAA-----TCAGTCAATGTAAACC  452
QY      256  AspAspGlyThrValTyrProGlnLysGlyArgLeuLeuPheAlaAspProValAlaAsn  275
Db      453  AATGCGATGGGTCTGT---CCGAAGCCACCATCTTCTACTGAACACAGATGAT  509
QY      276  GluSerThrGlnGlnIleThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeuMet  295
Db      510  GCTACTACCCGTTACCTTAAGTAAGAGCAGTATTG-----GATGAGCCCAATATTAAAT  563
QY      296  ProGlyLeuTyrValArgValLeuMetAsp  305
Db      564  CCGGTCCTTGTGTAAAGTATTATGAT  593

RESULT 28
LOCUS      B2558130      756 bp      DNA      linear      GSS 17-DEC-2002
DEFINITION pasci-60_808.s1 pasci-60 Pseudomonas aeruginosa genomic clone
ACCESSION  B2558130
VERSION    B2558130.1 GI:27171967
KEYWORDS   GSS.
SOURCE     Pseudomonas aeruginosa
ORGANISM   Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE  1 (bases 1 to 756)
            Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
            Whole-Genome-Sequence variation among multiple isolates of
            Pseudomonas aeruginosa library
            J. Bacteriol., (2002) In press
            Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source           1..756
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="1-60"
                     /db_xref="taxon:287"
                     /clone="pasci-60_808"
                     /clone_lib="pasci-60"
                     /note="clinical isolate 1-60 whole genomic shotgun
                     library."
BASE COUNT      140 a      263 c      223 g      130 t
ORIGIN
Alignment Scores:
Pred. No.:      1.18e-09      Length:      756
Score:          189.50      Matches:      67
Percent Similarity: 48.18%      Conservative: 39
Best Local Similarity: 30.45%      Mismatches: 77
Query Match:    9.39%      Indels:      37
DB:             29      Gaps:      9

US-09-889-756a-2 (1-412) x B2558130 (1-756)
QY      14  LeuAlaAlaValAlaLeuValLeuSerSerCysGly---LysGlyGlyAspAlaAla  32
Db      149  CTCGCCCGGTATGTCGTCGTGGGCGGATTCGCGCGGCTACAAAGCTTCACTCCATCCGT  208
QY      33  GlnGlyGlnProAlaGlnValProAlaProValValGlyValValThrVal  52

```

```

Db 209 CAG-----CAGATCGCCCTTTTACGACGCAAGAACCGCATGACGCTG----- 253
Qy 53 HisProGlnThrValAlaLeuThrValGluLeuPro----- 64
Db 254 -----ACCGCAGCTTGGCCGAAAGAGCTCGTGGAGAGCCGCTCCGACCATC 304
Qy 65 GlyArgLeuGluSerLeuArgThrAlaAspValArgAlaGlnValGlyIleLeuGln 84
Db 305 GCGACGCTCAAGCATTCACGAGCGCTGACCTCACCACCGAAGTCTCCGACGCTAGC 364
Qy 85 LysArgLeuPheGlnGlySerTyValArgAlaGlyGlnProLeuTyGlnIleAsp 104
Db 365 GACGTAAGTCTTCCGCGGACGACGAGTGAAGCTGACCAACCGCTATCAGATTGAA 424
Qy 105 SerSerThrTyGlnAlaAsnLeuGluSerAlaArgAlaGlnLeuAlaThrAlaGln 124
Db 425 AGCGACGTGAGAAAGCCACCTGCGC-----ACTGCGAGGCGC 463
Qy 125 ThrLeuAlaLysAlaAspAlaAspLeuAlaArgTyLysProLeuValAlaAlaGln 144
Db 464 GATCTCGGCGCTGCGAGGCGGAGTACGAGCGCGCGCACTGATCGGACGACAGGC 523
Qy 145 ValSerArgGlnGlyTyAsp-----AlaAlaValThrAlaLysArgSerAlaGln 162
Db 524 ATCTGAAAGCGAATTCGATCTCGCGCGCAGTGGCGCAAGACGACGCGCCACGCTC 583
Qy 163 GlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsnArgSer 182
Db 584 -----GCCGAGCTGAAAGCGCGC-----CTGCGCAAGAGCGC 616
Qy 183 IleThrAlaProIleSerGlyPheIleGlyGlnSerLysValSerGlyGlyThrLeu 202
Db 617 GTGCTCGGCGCTTCCGCGGACCATCCGCACTCCCGCAGTGCAGTGCAGCTAC 676
Qy 203 AsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAsnProMetTyVal 222
Db 677 TCGCCCGGG-----ACGCCGATCGCAACTTCGACGACCTTCACCTGCTCTGAT 730

RESULT 29
LOCUS B2340887 775 bp DNA linear GSS 24-APR-2001
DEFINITION B2340887 B. japonicum BAC library Bradyrhizobium japonicum
ACCESSION B2340887
VERSION B2340887.1 GI:13775936
KEYWORDS GSS
SOURCE Bradyrhizobium japonicum
ORGANISM Bradyrhizobium japonicum
REFERENCE Tomkins,J.P., Wood,T.C., Stacey,M.G., Itoh,T.T., Judd,A., Goicoechea
AUTHORS 'J.L., Stacey,G., Sadowsky,M.J. and Wing,R.A.
TITLE A marker-dense, sequence-ready map of the Bradyrhizobium japonicum
JOURNAL Genome Res. 11 (8), 1434-1440 (2001)
MEDLINE 21376150
PUBMED 11483585
COMMENT Contact: Wing RA
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Class: BAC ends
High quality sequence start: 8
High quality sequence stop: 711.
Location/Qualifiers
1..775
/organism="Bradyrhizobium japonicum"
/mol_type="genomic DNA"

```

```

BASE COUNT 168 a 250 c 215 g 139 t 3 others
ORIGIN
Alignment Scores:
Pred. No.: 1,76e-09 Length: 775
Score: 188.00 Matches: 58
Percent Similarity: 46.24% Conservative: 28
Best Local Similarity: 31.18% Mismatches: 89
Query Match: 9.31% Indels: 11
DB: 28 Gaps: 3
US-09-889-756a-2 (1-412) x B2340887 (1-775)
Qy 22 LeuSerSerGlyGlyGlyGlyAlaAlaGlnGlyGlnProAlaGlyArgGlu 41
Db 215 CTGCGCGAATCGCGGAAAGCCAGACGCGCAAAA-----AGCTCC 256
Qy 42 AlaPro-AlaProValValGlyValValThrValHisProGlnThrValAlaLeuThr 61
Db 257 GCCCGCGCGCGCGCGCTTCCGCGAGCGTTCGCGCGCGTCCAGAGTTCATTC 316
Qy 61 GluLeuPro-----GlyArgLeuGluSerLeuArgThrAlaAspValArgAlaGln 79
Db 317 CTACTTACCGGTCTCGGACGCTTACCGGTTTCAACACCTCCAGTTCGCGCGCGCT 376
Qy 79 GIGlyIleIleGlnLysArgLeuPheGlnGlySerTyValArgAlaGlnPro 99
Db 377 TGACGCGCAGATCGACAGATCGCTTACCGAAGCGCAGATCGTCAAGAGCGGAGCT 436
Qy 99 OleuTyGlnIleAspSerSerThrTyGlnAlaAsnLeuGluSerAlaArgAlaGln 119
Db 437 TCTGTCTCGATTCATCCCGCTCCCTATCAAGCGCGCTCGACGCGCAAGGCGAA 496
Qy 119 ValAlaThrAlaGlnAlaThrAlaLysAlaAspAlaAspLeuAlaArgTyLysPro 139
Db 497 GCGCGAGACGAGCAATTCGCGAAGCGCAAGCGCAATTCGAGTCCAGCGCGCATGA 556
Qy 139 ValAlaAlaGlnAlaValSerArgGlnGlyTyAspAlaAlaValThrAlaLys 159
Db 557 CGCGAATTCGCGACCGCGGCGGCGACACCGACCGGCTCCACTGCGCGCGGTTAC 616
Qy 159 ValGlnAlaGlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsn 179
Db 617 CGNCCAGATTCGC-----GCCGACCAAGCGCGGATCTCCATGCCGACCACTTGA 670
Qy 179 nargSerArgIleThrAlaProIleSerGlyPheIleGlyGlnSerLysValSerGlu 199
Db 671 CTACACCAAGTTCAGGAGGCGGATCCAGCGGCTGCGCGCGCGCGCGCGCGCG 730
Qy 199 YThrLeuLeuAla 204
Db 731 GAACATGTCTCAAGCC 746

RESULT 30
LOCUS B2340887 697 bp DNA linear GSS 06-NOV-2002
DEFINITION B2340887 B2340887 B. japonicum BAC library Bradyrhizobium japonicum
ACCESSION B2340887
VERSION B2340887.1 GI:24739302
KEYWORDS GSS
SOURCE Sorghum bicolor (sorghum)
ORGANISM Sorghum bicolor
REFERENCE Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
1 (bases 1 to 697)
Rabinowicz,P.D., O'Shaughnessy,A.L., Balija,V., Dedhia,N.,

```

TITLE
JOURNAL
COMMENT

Katzenburger, F., King, L., Miller, B., Muller, S., Nascimetto, L.,
Zutavern, T., Palmer, L., McCombie, W.R. and Mariensson, R.A.
Genomic shotgun sequences from *Sorghum bicolor* (methyl)-filtered
unpublished
Contact: W. Richard McCombie

```

Db      579  CGGCTACGTCGAGCGCNCNTGACGCGCGGTGACCGAAGCGGATCACCGTCCGCG 520
Oy      317  nglAlaValThrArgGlyAlaValAspThrValMet---lIeValaAlaGlnGly 336
Db      519  CGAGCGGTGATCCGTACCGCCAGTCCGCGGTGATGATGATCAACCAAGGGGCT 460
Oy      336  yMeGluProArgGluValThrValAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 356
Db      459  GGTGMAACGTGAGTCCGCGCGGACACCTCGAGGGCGGTGATGATGATGATGATGATG 400
Oy      356  rGlyLeuIleuAspGlyAspGlyValValValGlnGlnGlnGlnGlnGlnGlnGlnG 376
Db      399  CGGGCTCAAGGCGCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGAT 357
Oy      376  yAlaIleuValThrProGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 396
Db      356  -----CAGCATGCGCGCGGTGATGATGATGATGATGATGATGATGATGATGATG 316
Oy      396  nserGlyValGlnThrAlaSerGluAlaValThrAlaSerGluAla 411
Db      315  GCCGCGCAGCGCGCGCGCGCTACCGCTGCGCGGTGCGCGCG 270

RESULT 32
AZ302989/c 472 bp DNA linear GSS 06-MAR-2001
LOCUS      GSSBnu1931 Brucella abortus random genomic library Brucella
DEFINITION melitensis biovar Abortus genomic clone U01931, genomic survey
            sequence.
ACCESSION  AZ302989
VERSION     AZ302989
KEYWORDS    GSS.
SOURCE      Brucella melitensis biovar Abortus (Brucella abortus)
ORGANISM    Brucella melitensis biovar Abortus; Alphaproteobacteria; Rhizobiales;
            Bacteria; Proteobacteria;
REFERENCE   1 (bases 1 to 472)
            Sanchez,D.O., Zandomeni,R.O., Cravero,S., Verdun,R.E., Pierrou,E.,
            Faccio,P., Diaz,G., Lanzavecchia,S., Agüero,F., Frasch,A.C.C.,
            Anderson,S.G.E., Rosetti,O.L., Grau,O. and Ugaldé,R.A.,
            Gene discovery through genomic sequencing of Brucella abortus
            Infect. Immun. 69 (2), 865-868 (2001)
TITLE       JOURNAL MEDLINE
PUBMED     11159979
COMMENT     Contact: Siv Andersson
            Small Genomes Sequencing Group
            Department of Molecular Evolution, Uppsala University
            Nobdvägen 18C, S-752 36, Uppsala, Sweden
            Tel: 46-18-471-4379
            Fax: 46-18-471-6404
            Email: Siv.Andersson@bc.uu.se
            Sequences were basecalled with phred and vector was masked with
            crossmatch (see http://genome.washington.edu). Sequences were then
            trimmed from both ends to remove low quality bases and masked
            vector.
            Class: shotgun.
FEATURES
    source          location/Qualifiers
    1..472
    /organism="Brucella melitensis biovar Abortus"
    /mol_type="genomic DNA"
    /strain="2308"
    /db_xref="taxon:235"
    /clone="U01931"
    /clone_id="Brucella abortus random genomic library"
    /note="Vector: modified M13"

BASE COUNT      88 a 131 c 140 g 113 t
ORIGIN

```

```

Query Match:      8.94%      Indels:      11
DB:               28      Gaps:           3

US-09-889-756a-2 (1-412) x AZ302989 (1-472)

Oy      249  lIeAlaValGlyIleuValPheAspAspGlyThrValIleProGlnGlyValGluLeu 268
Db      466  GTAGCGGTGATCCGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 407
Oy      269  PheAlaAspProValValaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 288
Db      406  TTATCCGGATCAAGATTAAATCCCAATACGATGATGATGATGATGATGATGATGAT 347
Oy      289  AsnAspGlnAsnIleLeuMetProGlyLeuThrValAlaValLeuMetAspGlnValAla 308
Db      346  AATCCGAACAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 287
Oy      309  ValAspAsnAlaPheValValProGlnGlnAlaVal---ThrArgIleValAspThr 327
Db      286  CCGGTGAACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 227
Oy      328  ValMetIleValAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 347
Db      226  GTTCTGCGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 167
Oy      348  GlnGlyThrAsnThrIleValThrSerGlyLeuValAspGlyAspValValValGlu 367
Db      166  TTTCGACGACATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 107
Oy      368  GlyIleSerIleAlaGlyIleThrGlyAlaValValValValValValValValVal 380
Db      106  GGCTTCAG-----CCGATCCGCGCGCGCGAGAGGTGATGATGATGATGATGATGAT 53
Oy      381  ---ThrProGlyGlnThrAlaSerSer 388
Db      52  ACCAGCGCGCGCGCGAGAAAGCATGCTG 26

RESULT 33
BH247218/c 450 bp DNA linear GSS 26-NOV-2001
LOCUS      BOGAB20TR Brassica oleracea genomic clone BOGAB20, genomic
DEFINITION survey sequence.
ACCESSION  BH247218
VERSION     BH247218
KEYWORDS    GSS.
SOURCE      Brassica oleracea
ORGANISM    Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; euroids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1 (bases 1 to 450)
            Town,C.D., Van Aken,S., Uteirback,T., Koo,H. and Fraser,C.M.
            Whole genome shotgun sequencing of Brassica oleracea
            unpublished
            Other GSSs: BOGAB20TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES
    source          location/Qualifiers
    1..450
    /organism="Brassica oleracea"
    /mol_type="genomic DNA"
    /strain="TO1000DH3"
    /db_xref="taxon:3712"
    /clone="BOGAB20"
    /clone_id="BOGA"
    /note="Vector: pHD01, Site_1: BseXI; 2-3 kb sheared

```

Alignment Scores:

Pred. No.:	5.52e-09	Length:	472
Score:	180.50	Matches:	48
Percent Similarity:	51.01%	Conservative:	28
Best Local Similarity:	32.21%	Mismatches:	62

BASE COUNT 57 a 151 c 158 g 84 t
 ORIGIN genomic DNA inserted into pHO51 using BstXI linkers"

Alignment Scores:

Pred. No.: 6,596-09 Length: 450
 Score: 179.50 Matches: 51
 Percent Similarity: 46.67% Conservative: 19
 Best Local Similarity: 34.00% Mismatches: 65
 Query Match: 8.89% Indels: 15
 Gaps: 1

US-09-889-756a-2 (1-412) x BH247218 (1-450)

QY 43 ProAlaProValValGlyValValThraValHisProGlnThrValAlaLeuThrValGlu 62
 Db 428 CCCGCTCTTCAACCGCGGTGCGCCCTCAACGCCGAG----- 390
 QY 63 LeuProGlyArgLeuGluSerLeuArgThrAlaAspValArgAlaGlnValGlyTle 82
 Db 389 -----TACGACGTGACCGTGAAGAGCCAGCTGACGCCAG 354
 QY 83 IleGlnlyArgLeuPheGlnGluGlySerTyrValArgAlaGlnProLeuTyrGln 102
 Db 353 ATCAACCATGTGCGGTTCACAGAGCCAGAGGTGCGCGGCGAGCTGCTGCGAG 294
 QY 103 IleAspSerSerThrTyrGluAlaAsnLeuGluSerAlaArgAlaGlnLeuAlaThrAla 122
 Db 293 ATTCATCGCCCGCGGCTCCAGGCCAGCCAGCCAGCGAGCGAGCTGCTGCGAGC 234
 QY 123 GlnAlaThrLeuAlaValAspAlaAspLeuAlaArgTyrIleProLeuValAlaAla 142
 Db 233 AAGCGAGCGCTCGCCCAAGCGCGGCTCGAGCTCGCGCCATCGAGCGGCGGAAATG 174
 QY 143 GlnAlaValSerArgGlnGluTyrAspAlaAlaValThrAlaValArgSerAlaGluAla 162
 Db 173 AACCGCGCGCGGTGCGAGTCTGATCATGAAAGCGCGCTCAACGAGCTGCATGCG 114
 QY 163 GlyValIysAlaAlaGlnAlaAlaIleYserAlaGlyIleAsnLeuAsnArgSerArg 182
 Db 113 CAGATCCGCGCGCATCAGCGCGCGCTGCAGAACCGCCGCGCTGCTGATTAACAGACG 54
 QY 183 IleThrAlaProIleSerGlyPheIleGly 192
 Db 53 ATCCGCGCGCCATCTCGCGGCGCATCGCC 24

RESULT 34
 BH242325/c 745 bp DNA linear GSS 13-NOV-2001
 LOCUS ATZFB03TR ATZF Arabidopsis thaliana genomic clone ATZFB03, genomic
 DEFINITION survey sequence.

ACCESSION BH242325
 VERSION BH242325.1 GI:16917423
 SOURCE GSS.

ORGANISM Arabidopsis thaliana (thale cress)
 SOURCE Arabidopsis thaliana

REFERENCE 1 (bases 1 to 745)
 AUTHORS Kaul, S., Town, C.D., Bowman, C.L., Van Aken, S.E., Utterback, T.V.,
 Feldblyum, T.V. and Fraser, C.M.
 Survey sequencing of Arabidopsis thaliana BAC T20G10

TITLE Unpublished
 JOURNAL Other GSSs: ATZFB03TF
 COMMENT Contact: Chris Town

7912 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@ctigr.org
 From Wash. U contig 1370. Caution: the DNA in this BAC may be from
 some non-Arabidopsis source

Seq primer: TR
 Class: sheared ends.
 Location/Qualifiers

FEATURES

1..745
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="ATZFB03"
 /clone_lib="ATZFP"
 /note="Vector: pHO52, Site 1: BstXI, 2-3 kb sheared BAC
 DNA inserted into pHO52 using BstXI linkers"

BASE COUNT 136 a 211 c 233 g 163 t
 ORIGIN 2 others

Alignment Scores:

Pred. No.: 2,416-08 Length: 745
 Score: 177.00 Matches: 57
 Percent Similarity: 38.81% Conservative: 28
 Best Local Similarity: 26.03% Mismatches: 68
 Query Match: 8.77% Indels: 66
 Gaps: 4

US-09-889-756a-2 (1-412) x BH242325 (1-745)

QY 75 ValArgAlaGlnValGlyGlyIleIleGlnlyArgLeuPheGlnGluGlySerTyrVal 94
 Db 663 GTCCGCGGAAAGTAAAGATATATACAGACCGCTCGCGATCGACGACACACCGCTC 604
 QY 95 ArgAlaGlyGlnProLeuTyrGlnIleAspSerSerThrTyrGluAlaAsnLeuGluSer 114
 Db 603 CGAGCCGCGACGCTTGTTCGCGATCATGACCGGATTAATCGCGAAATTAAGAA 544
 QY 115 AlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaValAspAlaAsp----- 132
 Db 543 GCGGTGCAATCTGAAGCGCCAGAGCTGCTTCCATCATGTCAGCGGAGACGAA 484
 QY 132 ----- 132
 Db 483 CTTCACACGACCTATTTCGCGAGCTGAGGCTCTACGGGCTCTCGGAACCGCATTC 424
 QY 133 --LeuAla-----ArgTyrIleProLeuValAlaAlaGluAlaValSer 146
 Db 423 AGCTGCGCATCAAGCGTCCGACCGCGCGAGCTCATTTCCAGCAACGCGCTCAGT 364
 QY 147 ArgGlnIleTyrAspAlaAlaValThrAlaValArgSerAlaGluAlaGlyValAla 166
 Db 363 CAAGCCGAGCTCATGAAACCGATACAAAGCATTCGAAACCGCGGCTGTATCGGCC 304
 QY 167 AlaGlnAlaAlaIleYser----- 173
 Db 303 GCATCAGCAGCGCTCGACTCGCAACGCTTGCGCTTCCGTTATGCGCGCAACGTAA 244
 QY 174 -----AlaGlyIleAsnLeu 178
 Db 243 GCTGCCCTCGCGCGATGCTGCTGAGGCTGAAGCGCCCGCGAGATCGACAGATGATCG 184
 QY 179 AsnArgSerArgIleThrAlaProIleSerGlyPheIleGlnlySerIysValSerGlu 198
 Db 183 GACAAATACGTTTGTGACAGCGCGCGTGTGCGGGGTATCGGAAACCGCGAGTTCCGATA 124
 QY 199 GlyThrLeuLeuAsnAlaGly-----AspThrThr 208
 Db 123 GGTGAGCTTGTCCACCGGCTGCTCTTCTGCTGATATCTCCCGGTGAGCGAGTGTGG 64
 QY 209 ValLeuAlaThrIleArgGlnThrAsnProMetTyrValAsnValThrGlnSerAla 227
 Db 63 ATTGTGCAATTTCAAGAAACCGAGTCCGACATCAAGATCGGTCAACAGCC 7
 RESULT 35
 B2562720 847 bp DNA linear GSS 17-DEC-2002
 LOCUS B2562720
 DEFINITION pacs2-164_3895.y2 pacs2-164 Pseudomonas aeruginosa genomic clone

ACCESSION	pac2-164_3895, genomic survey sequence.					
VERSION	BZ562720					
KEYWORDS	BZ562720.1 GI:27185220					
SOURCE	GSS.					
ORGANISM	<i>Pseudomonas aeruginosa</i>					
REFERENCE	<i>Pseudomonas aeruginosa</i>					
AUTHORS	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.					
TITLE	1 (bases 1 to 847)					
JOURNAL	Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,T.L., Kaul,R. and Olsen,M.V.					
COMMENT	Whole-genome-sequence variation among multiple isolates of <i>Pseudomonas aeruginosa</i> library J. Bacteriol., (2002) In press					
	Contact: Chris K. Raymond					
	Genome Center					
	University of Washington					
	Box 352145, Seattle, WA 98105-2145, USA					
	Tel: 2062216954					
	Fax: 2066857244					
	Email: ckraymond@u.washington.edu					
FEATURES	Class: shotgun.					
source	Location/Qualifiers					
	1..847					
	/organism="Pseudomonas aeruginosa"					
	/mol_type="genomic DNA"					
	/strain="2-164"					
	/db_xref="taxon:287"					
	/clone="pac2-164_3895"					
	/clone_1b="pac2-164"					
	/note="clinical isolate 2-164 Whole genomic shotgun library."					
BASE COUNT	196 a	269 c	236 g	146 t		
ORIGIN						
Alignment Scores:						
Pred. No.:	6.7e-08	Length:	847			
Score:	173.50	Matches:	61			
Percent Similarity:	43.06%	Conservative:	32			
Best Local Similarity:	28.24%	Mismatches:	86			
Query Match:	8.59%	Indels:	37			
DB:	29	Gaps:	6			
US-09-889-756A-2 (1-412) x BZ562720 (1-847)						
Cy	96	AlAGlYgInProLeuTYrGlInleAsPSerSerThTYrGlUnlaSnleUgInserAla	115			
	:::::	:::::	:::::			
Db	1	AGCGGCCTTCCTCTACTAGATCAAGAACTGGACCTCCACC	57			
	:::::	:::::	:::::			
Cy	116	ArgAlaGlInleuAlaThrAlaGlInlaTrnleuAlaTyAlaSPalaSPleuAlaArg	135			
	:::::	:::::	:::::			
Db	58	CTAGAACACTGTGATCCCCCTGGGGCACTAAGAACGCCAGATCAGCTGCACGCC	117			
	:::::	:::::	:::::			
Cy	136	TyrLYSProLeuValAlaIaGluaValaSerArgIngluTYrASpaLaalAvalThr	155			
	:::::	:::::	:::::			
Db	118	TACAAGGGCGCTGTAGCCGAGGACTGCATGCGCAAGACAGCCTGGATTACCAGGAACC	177			
	:::::	:::::	:::::			
Cy	156	AlaLyshArgSerAlaGlualaGlyValLySaAlaAlaGlInlaAlaIleYsserAlaGly	175			
	:::::	:::::	:::::			
Db	178	CAGGTCGCCGACGTGGACGGACCACTCCGTACCAACAGGGCCAGGTCCAGCACGCCGC	237			
	:::::	:::::	:::::			
Cy	176	IleasnleuAnslargSerArgIleThrAlaProIleSerGlyhetleglyGlnserLys	195			
	:::::	:::::	:::::			
Db	238	CTCAACCTTACCTTCAACGAGGTCCGGACCATTTCCGAGGCGCTCGGCTCGCCAG	297			
	:::::	:::::	:::::			
Cy	196	ValsergluglyThrleuDeusnaIagIyaSPThrThValleuAlaThrIIeaArgIn	215			
	:::::	:::::	:::::			
Db	298	GTGACATCGGCAACTCGGTACACAGCGGAGATACACAGCGCGGTGGTGATCACCCAG	357			
	:::::	:::::	:::::			
Cy	216	ThraenPrometTYrValasnValThrInserAlaSerGluValmetylsleuAgArg	235			
	:::::	:::::	:::::			
Db	358	GTCAGGCCATC-----TCGGTGGGTTCACGCTGCCGCA	393			

Oy	236	Gnrlleatlaagluclyllysleutendulaalaaspglyvalillealvaliclylleuspe	255
Dd	394	CAG-----	396
Oy	256	AspaapgylythrValTyr-----ProglulygylArgleuleupeha	270
Dd	397	CAGATCGGCACCCGTGTCGAGCATGAACGGCCCCGCCAAGTGACGCTCACCCGCTG	456
Oy	271	AaspProAlvalAsnGluserThr-----GlyGlnIleThrleukrglaAlaIval	287
Dd	457	GACCCGAACCAAGAACAAAGGTTCTTGCCGAAGAAGACCCTTAACAACCCCTTGACAAAC	516
Oy	288	ProAnaapGlnAsnIleleuwerProglyleutyryValArgValIleu	303
Dd	517	CCA---GATTTCGAACACCACTACTCCGGGAAACGGGTCAAAGCTT	561
RESULT 36	BZ549534	793 bp	DNA linear GSS 17-DEC-2000
LOCUS	pacsl-60_2148.x1 pacsl-60	Pseudomonas aeruginosa genomic clone	
DEFINITION	pacsl-60_2148, genomic survey sequence.		
ACCESSION	BZ549534		
VERSION	BZ549534.1	GI:27153115	
KEYWORDS	GSS.		
SOURCE	Pseudomonas aeruginosa		
ORGANISM	Pseudomonas aeruginosa		
REFERENCE	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;		
AUTHORS	Pseudomonadaceae; Pseudomonas. 1 (bases 1 to 793) Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V. Whole-genome-sequence variation among multiple isolates of Pseudomonas aeruginosa library J. Bacteriol., (2002) In press Contact: Chris K. Raymond Genome Center University of Washington Box 352145, Seattle, WA 98105-2145, USA Tel: 20622316954 Fax: 2066857244 Email: ckraymond@u.washington.edu Class: Shotgun. Location/Qualifiers 1..793 /organism="Pseudomonas aeruginosa" /mol_type="genomic DNA" /strain="1-60" /db_xref="taxon:287" /clone="pacsl-60_2148" /clone_lib="pacsl-60" /note="Clinical isolate 1-60 whole genomic shotgun library."		
FEATURES	source		
BASE COUNT	146 a	245 c	250 g 151 t 1 others
ORIGIN			
Pred. No.:	7	8e-08	length: 793
Score:	172.50	Matches:	51
Percent Similarity:	50.00%	Conservative:	18
Best Local Similarity:	36.96%	Mismatches:	64
Query Match:	8.54%	Indels:	5
DB:	29	Gaps:	2
US-09-889-756A-2 (1-412)	x BZ549534 (1-793)		
Oy	16	AlaAlaValAlaLeuValleuSerCysGlyLySGlyGlyASPAlaAlaGlnGlyGly	35
Dd	437	GGGGGGCCCGCCTCGCGCGCTGCTCTGACGGGGCAACTCGGT-----CNCGCGCGC	387
Oy	36	GlnProAlaGlyArg-----GlnAlaProAlaProValAlaGlyValAlaThrValHis	53
Dd	386	AATTCGGCGCGCGCGCTCCCAAGGCAACAGCGCTCACCGTGGCGTGGCCAGGCGTAG	327

QY 54 ProGlnThrValAlaLeuThrValGluLeuProGlyArgLeuGluSerLeuArgThrAla 73
 Db 326 CAGGCGCAGCTGGCGGTCATTCACACCGCTTCCACCGCTTCCACACCGGTTG 267
 QY 74 AspValArgAlaGlnValGlyGlyLeuLeuGlnArgLeuPheGlnGlnGlySerTyr 93
 Db 266 AACGTCACAGCCGCGGTCACGCGAGCTGTCAAGGTGCTTCCAGAGGCGGCGAGAG 207
 QY 94 ValArgAlaGlnGlnProLeuTyrGlnLeuAspSerSerThrTyrGlnAlaAsnLeuGlu 113
 Db 206 GTCAAGCGCGCGACCTGCTGGCGGTGTCACCGCGACCTTCCACAGCGCGCGCTGCC 147
 QY 114 SerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaTyrAlaAspAlaAspLeu 133
 Db 146 CAGCGCCAGGCGCAGCTGATACAGATCCAGCGCAGCACTGAAGAACAACAAGATGACGTG 87
 QY 134 AlaArgTyrLysProLeuValAlaAlaGlnAlaValSerArgGlnGlyTyrAsp 151
 Db 86 CAGCGGCAACAAGGTGTGTATGCGCTGCAATGCTTCCGCGGCGAGCGCGGAT 33

RESULT 37
 BZ638338/c 876 bp DNA linear GSS 29-JAN-2003
 LOCUS OGCGN80TC_ZM_0.7_1.5_KB_Zea_mays genomic clone ZMMBHa0141M16,
 DEFINITION genomic survey sequence.
 ACCESSION BZ638338
 VERSION BZ638338.1 GI:28090523
 KEYWORDS GSS.
 SOURCE Zea mays
 ORGANISM Zea mays

REFERENCE
 AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick
 'A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Kohlfing,T., Citek
 'R.W., Numberg,A., Robbins,D. and Lakey,N.
 TITLE Consortium for Maize Genomics
 JOURNAL Unpublished
 COMMENT Other GSSs: OGCGN80TM
 CONTACT: Cathy Whitelaw
 TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-838-5843
 Fax: 301-838-0208
 Email: whitelaw@tigr.org
 Seq primer: TF
 Class: sheared ends.

FEATURES
 Source
 1. 876
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /db_xref="taxon:4577"
 /clone="ZMMBHa0141M16"
 /clone_1kb="ZM_0.7_1.5_KB"
 /note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
 methylation filtered genomic DNA library"

BASE COUNT 114 a 287 c 326 g 149 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1.45e-07 Length: 876
 Score: 170.50 Matches: 82
 Percent Similarity: 40.58% Conservative: 43
 Best Local Similarity: 26.62% Mismatches: 120
 Query Match: 8.44% Indels: 63
 DB: 29 Gaps: 12

US-09-889-756a-2 (1-412) x BZ638338 (1-876)

QY 97 GlyGlnProLeuTyrGlnLeuAspSerSerThrTyr----- 108
 |||||
 |||||

Db 856 GCGCAGCCAAAGCTCTTCATCAAGACGAGCATTTGTCGATGCCGCTGCTTCT 797
 QY 109 ---GlnAlaAsnLeu-GluSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAl 127
 Db 796 CCGCGCGCGCGGCTACCAAGACCGCGAGCGCGAGCTGTGCGCGCGAG-----CG 746
 QY 127 AluValaAspAlaAspLeuAlaArgTyrLysProLeuValAla-AlaGlnAlaValaSer 147
 Db 745 CAAGCGCGACCGCAGCAGCAAAATATATAGACCGCGCGCTGCGCGGTGAAGATATCAGC 686
 QY 147 rgtGlnGlyTyrAsp-----AlaAlaValThrAlaLysArgSerAlaGlnAlaGlyVal 165
 Db 685 AGCCCAAGGCGGACCTCGCGCGCGCGAGCGACCGCGCGCGCGCGCGCGCTGCTG 626
 QY 165 YsaIaAlaGlnAlaAlaIle-----LysSerAlaGlyLysLeuAsnLysArg- 180
 Db 625 GCGCGCGCGCGCAAGCTGCGATCTGTGGGCAAGTCCAGGCG--GAGATCAACCGC 569
 QY 181 -----SerArgIleThrAlaProIles 188
 Db 568 TGGAGATGTCGCGCAAGTCTCCGATCCAGCATCACAGCATCCAGCGTCCAGCACCGATCA 509
 QY 188 erGlyPheIleGlyGlnSerLysValSerGlnGlyThrLeuAsnAlaGlyAspThr 208
 Db 508 GCGCGCTGCTGCGCTGCGCGGATGTCGCGCGGCAATATGTGTGCGCAGGCGCGCA 449
 QY 208 hrValLeuAlaThrIleArgGlnThrAspProMetTyrValaAsnValThrGlnSerAla 228
 Db 448 AGCCGTGATGACCATC-----ACCGATCCGTGCGCGCTA----- 414
 QY 228 erGluValMetLysLeuArgArgGlnIleAlaGlnGlyLysLeuAlaAlaAspGly 248
 Db 413 -----TGCCTGTCGCGCGAGTGGCGGAGCGCAT----- 384
 QY 248 allAlaValAlaGlyIleLysPheAspAspGly-----ThrValTyrProGlnLysG 265
 Db 383 -----GCCGAGATGTCGCGTGGCGGATCCGTCGAGTGAAGACCCCGCGCTGCGG 329
 QY 265 lYArgLeuLeuPheAla-----AspProValValaGlnGluSerThrGlyG 280
 Db 328 GCCGCTCTTCCAGCGGAGCATGATGTCGCGCGCGCGCGTGAACCGCGAGCGAC 289
 QY 280 lntLeuThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetProGlyLeuTyrV 300
 Db 268 GCGTGGCGGTCGCGCTGATCCCAACCGCGAGCGCGTGAAGCCCAATGTCG 209
 QY 300 aArgValLeuMetAspGlnValAlaValaAspAsnAlaPheValValProGlnGlnAla 320
 Db 208 CCAGCTTCGCGATCAAGCATCTGAACGCGGCGAAGCGATCCGCGCGCGCGCG 149
 QY 320 aLThr---ArgGlyAlaLysAspThrValMetLeuValaAsnAlaGlnGlyMetGluP 339
 Db 148 TCATTCATGAAGCGACACCGCGCGCTGTGATCTAGCGCGCGACCGCGCTGCTGTCG 89
 QY 339 roArgGlnValThrValAlaGlnGlnGlnGlyThrAsnTrpIleValThrSerGlyLeu 359
 Db 88 CCGCGCAGCTGAACGAGGCGGACTCCGCCAATGCTGTGTCACCATCACCAGGCGCTGA 29
 QY 359 YsaIAspGlyAspLysValVal 365
 Db 28 AGACCGCGCGAAGATGCTC 9

RESULT 38
 BP416737 469 bp mRNA linear EST 28-NOV-2000
 LOCUS UI-R-CAO-bku-g-12-0-UI.s1 UI-R-CAO Rattus norvegicus cDNA clone
 DEFINITION UI-R-CAO-bku-g-12-0-UI 3', mRNA sequence.
 ACCESSION BP416737
 VERSION BP416737.1 GI:11404688
 KEYWORDS EST.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;


```

Db 191 GAAGGCGAGATGTCGAGGCGGCGAGTTGCTGGCGACCATGATGACCGAGCGAGTTC 250
Qy 110 AlaAsnLeuGluSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaAla 129
Db 251 CTCGGCGTGGACGACCGCGAGCCCTCCAGAGTACGAGCCAGCCAGCTGAAATCCGAC 310
Qy 130 AspAlaAspLeuAlaArgTyrLeuProLeuValAlaAlaGlnAlaValSerArgGln 149
Db 311 GAGCAGAGACCTGCAACGCTACCGAGCTGTTGGCCGAGCGTGGTAACTCTACTG 370
Qy 150 TyrAspAlaAlaValThrAlaLysSerAlaGlnAlaGlyValLysAlaAlaGlnAla 169
Db 371 TTGATTCAGCAACAGCGACGCTCAGACGCTGCGTGGACCTTAATGCCAAGATGCC 430
Qy 170 AlaIleLysSerAlaGlyIleAsnLeuAsnArgSerArgIleThrAlaProIleSerGly 189
Db 431 ATCATCAACCCCGACGACGCTGGCTGTATCACTGCGATCAGCTCGTCCGCTCGCGT 490
Qy 190 PheIleGlyGlnSerLysValSerGlnGlyThrLeuLeuAsnAlaGlyAspThrVal 209
Db 491 AGGCGCGGTATCCGAAACGTCATGTCGAAACGCGAGCGGTCGTCGACAGCTTCTCT 550
Qy 210 LeuAlaThrIleArgGlnThrAspProMetTyrValAsnValThrGlnSerAlaSerGlu 229
Db 551 CTGTTCAAGATGCGACCAATCGATCGATGAGTCTTCTCTCACTGCAACAGAAACAG 610
Qy 230 ValMetLysLeuArgArgGlnIleAlaGlnGlyLysLeuLeuAlaAlaAspGlyValIle 249
Db 611 TTGCCCCAGTTGC--AGGCACCTGCTCGGCGGGAAGCGGAGTGCACCCCTACAGTCCA 668
Qy 250 AlaValGlyIleLysPheAspAspGlyThrValTyrProGlnLysGlyArgLeuLeuPhe 269
Db 669 ACAGTTGAGGCTCCCTCCACAGAACGCCGTTGCTTGCACATCG-- 713
Qy 270 AlaAspProValValAsnGluSerThrGlyGlnIleThrLeuArgAlaAlaValPro--- 288
Db 714 ---CAACCAATCGCGAGCTTCAACCGGA-----ACCTTCGCGGAGCGGCCCTTC 764
Qy 289 -----AsnAspGlnAsnIleLeuMetProGlyLeu 298
Db 765 CTTATATTAACNCGCGAGGCGCGGCTTCTTGGCCCGGCTCA 806

RESULT 40
BZ578763 1228 bp DNA linear GSS 17-DEC-2002
LOCUS msh2.5991.y2 msh Pseudomonas aeruginosa genomic clone msh2_5991,
DEFINITION Genomic survey sequence.
ACCESSION BZ578763
VERSION BZ578763.1 GI:27213824
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 1228)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence Variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL U. Bacteriol., (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES
source 1. 1228
location/Qualifiers
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="MSh"

```

```

/db_xref="taxon:287"
/clone="msh2_5991"
/clone_id="msh"
/notes="Environmental isolate. Whole genomic shotgun
library."
BASE COUNT 226 a 401 c 35t g 242 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 1,26e-06 Length: 1228
Score: 163.50 Matches: 64
Percent Similarity: 43.75% Conservative: 41
Best Local Similarity: 26.67% Mismatches: 100
Query Match: 8.10% Indels: 36
DB: 29 Gaps: 6
US-09-889-756a-2 (1-412) x BZ578763 (1-1228)
Qy 8 AlaMetArgAlaAlaAlaLeuAlaAla-----AlaValAlaLeuValLeuSerSerCys 25
Db 206 GCCATGAAGCGGTTTCTCTCGCGCGCTGTCGCTTCCCTCTCTCTCTCTCTCTCTCT 265
Qy 26 GlyLysGlyLysAspAlaAlaGlnGlyGlnProAlaGlyArgGlnAlaProAlaPro 45
Db 266 GGC-----GACGAGCGCGCGGACCG 289
Qy 46 ValValGlyValValThrValHisProGlnThrVal-----AlaLeu 59
Db 290 CCGCGCGCGGCTGCTGACGCGTGAAGACCTGTAAGAACGACGACCTCGGCTTC 349
Qy 60 ThrValGlnLeuProGlyArgLeuGlnSerLeuArgThrAlaAspValArgAlaGlnVal 79
Db 350 GCGGAGCATTCACAGGCGCGCTACGAGAGCGTG-----CTCGGCTTCCGAC 397
Qy 80 GlyGlyIleIleGlnLysArgLeuPheGlnGlnGlySerTyrValArgAlaGlyGlnPro 99
Db 398 AACGAGCGATTCGCTCGCTCGCTGCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTG 457
Qy 100 LeuTyrGlnIleAspSerSerThrTyrGlnAlaAsnLeuGluSerAlaArgAlaGlnLeu 119
Db 458 CTGGCGACCTTCGACCCACGACGACGACGACGACGACGACGACGACGACGACGACGAC 517
Qy 120 AlaThrAlaGlnAlaThrLeuAlaLysAlaAspAlaAspLeuAlaArgTyrLysProLeu 139
Db 518 GCCAGCGCGAGGACAGTGTATGATGACGCGCCATGCGCAATGCCGCGCGCGAGAAAGCTG 577
Qy 140 ValAlaAlaGlnAlaValSerArgGlnGlnTyrAspAlaAlaValThrAlaLysArgSer 159
Db 578 TTGCTCGCAGCGTCAACCGCGCGCGCGCTGAGCATGCGGAGCCGCTGTAACACC 637
Qy 160 AlaGlnAlaGlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsn 179
Db 638 AGCCAGCGCAGCTTCGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 697
Qy 180 ArgSerArgIleThrAlaProIleSerGlyPheIleGlyGlnSerLysValSerGly 199
Db 698 TACACGCGCTGTGACCATTTGACGCGGCTTAT----- 733
Qy 200 ThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAspProMet 219
Db 734 ACCACCTGTCACCGCGAGCGCGCATGTGTTGACGCGCGCGCGCGGCTGTTACCTTTG 793
Qy 220 TyrValAsnValThrGlnSerAlaSerGlnValMetLysLeuArgArgGlnIleAlaGln 239
Db 794 CCCGTC-----CGAGTGGCAGCAAGTCTTTGA-CTGGCCAAACGAGTTGCCGAG 843

```

Search completed: September 8, 2003, 07:19:43
Job time : 2763 sec

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 8, 2003, 07:07:31 ; Search time 343 Seconds
(without alignments)
2762.655 Million cell updates/sec

Title: US-09-889-756A-2
Perfect score: 2019
Sequence: 1 MAFYAFKAMRAAAALAAVAL.....APOSQVOTSEAKTASEAE 412

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1537136 seqs, 1149988732 residues
Total number of hits satisfying chosen parameters: 3074272

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n model -DEV=xlh
-Q=/cgnr1/USPFO.apool/US09889756/runat_05092003_124331_23437/afp_query.fasta_1.583
-DB=Published Applications NA -OFMT=fastap -SUFFIX=tmpb -MINMATCH=0.1
-LOOPCL=0 -LOOEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=bl0sum62
-TRANS=numan40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=40 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09889756 @CGN 1 1 244 @runat_05092003_124331_23437
-NCPU=6 -ICPU=3 -NO MMAP -LARGEBUERY -NEG_SCORES=0 -WAIT -DSPELOCK=100
-LONLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:
1: /cgnr2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
2: /cgnr2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:
3: /cgnr2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:
4: /cgnr2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:
5: /cgnr2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:
6: /cgnr2_6/ptodata/2/pubpna/PCUUS_PUBCOMB.seq:
7: /cgnr2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:
8: /cgnr2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:
9: /cgnr2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:
10: /cgnr2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:
11: /cgnr2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:
12: /cgnr2_6/ptodata/2/pubpna/US09 NEW_PUB.seq:
13: /cgnr2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:
14: /cgnr2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:
15: /cgnr2_6/ptodata/2/pubpna/US10 NEW_PUB.seq:
16: /cgnr2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:
17: /cgnr2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match Length	ID	Description
C 1	701.5	34.7	7886	14	US-10-114-170-102

2	435.5	21.6	1395	9	US-09-912-020-121	Sequence 121, App
3	370.5	18.4	1008	11	US-09-975-719-148	Sequence 148, App
4	286	14.2	1830121	14	US-10-329-960-1	Sequence 1, Appl
5	200	9.9	1512	11	US-09-975-719-171	Sequence 171, App
6	175	8.7	1977	10	US-09-738-626-3315	Sequence 3315, App
7	175	8.7	3309400	10	US-09-738-626-1	Sequence 1, Appl
8	173.5	8.6	930	9	US-09-815-242-9752	Sequence 9752, Ap
9	169	8.4	900	9	US-09-815-242-6071	Sequence 6071, Ap
10	162.5	8.0	858	9	US-09-815-242-7815	Sequence 7815, Ap
11	159.5	7.9	1830121	14	US-10-329-960-1	Sequence 1, Appl
12	156.5	7.8	1173	9	US-09-815-242-7049	Sequence 7049, Ap
13	151.5	7.5	1200	12	US-09-769-744A-93	Sequence 93, Appl
14	140.5	7.0	9025608	14	US-10-156-761-1	Sequence 1, Appl
15	138.5	6.9	1068	9	US-09-741-669-242	Sequence 242, App
16	134.5	6.7	3000	10	US-09-855-154-1	Sequence 1, Appl
17	133.5	6.7	3000	9	US-10-327-353-3	Sequence 3, Appl
18	133.5	6.6	4556	14	US-09-764-208-1	Sequence 1, Appl
19	133.5	6.6	4556	14	US-10-315-023-1	Sequence 1, Appl
20	131.5	6.5	2862	14	US-10-156-761-2565	Sequence 2565, Ap
21	131.5	6.5	3000	14	US-10-227-353-5	Sequence 5, Appl
22	131.5	6.5	3116	10	US-09-855-754-3	Sequence 3, Appl
23	130.5	6.5	2398	14	US-10-114-170-260	Sequence 260, App
24	130.5	6.5	25165	14	US-10-114-170-39	Sequence 39, Appl
25	129.5	6.4	1539	10	US-09-938-842A-176	Sequence 176, App
26	128.5	6.4	9025608	14	US-10-156-761-1	Sequence 1, Appl
27	127.5	6.3	1116	11	US-09-988-067B-51	Sequence 15, App
28	125.5	6.2	1794	9	US-09-793-306-145	Sequence 145, App
29	125	6.2	105164	11	US-09-847-513A-1	Sequence 1, Appl
30	122.5	6.1	990	8	US-08-945-038-3	Sequence 3, Appl
31	122.5	6.1	5505	14	US-10-156-761-413	Sequence 413, App
32	122.5	6.1	100000	14	US-10-156-761-15103	Sequence 15103, A
33	121.5	6.0	12145	11	US-09-308-207-19	Sequence 19, Appl
34	121.5	6.0	12145	12	US-10-277-249-1	Sequence 1, Appl
35	121.5	6.0	12145	14	US-10-213-203-1	Sequence 1, Appl
36	120.5	6.0	3543	12	US-10-193-764-60	Sequence 60, Appl
37	120.5	6.0	3568	12	US-10-193-764-58	Sequence 58, Appl
38	119.5	5.9	2088	14	US-10-156-761-1788	Sequence 1788, Ap
39	119.5	5.9	3234	12	US-10-190-435-51	Sequence 51, Appl
40	119.5	5.9	46819	14	US-10-114-170-72	Sequence 72, Appl
41	119	5.9	1116	14	US-10-156-761-5046	Sequence 5046, Ap
42	119	5.9	3695	14	US-10-114-170-18	Sequence 18, Appl
43	118	5.8	2733	10	US-09-855-754-2	Sequence 2, Appl
44	118	5.8	3353	11	US-09-927-827-41	Sequence 41, Appl
45	118	5.8	7755	10	US-09-971-536-26	Sequence 26, Appl

ALIGNMENTS

RESULT 1
US-10-114-170-102/c
Publication No. US20030023075A1
GENERAL INFORMATION:
APPLICANT: Blatner, Frederick R.
Burland, Valerie
Perna, Nicole T.
Plunkett, Guy
Welch, Rod
TITLE OF INVENTION: No. US20030023075A1 Sequences of E. coli O157
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0
CURRENT APPLICATION DATA:

```

1      APPLICATION NUMBER: US-10/114,170
2      FILING DATE: 01-Apr-2002
3      CLASSIFICATION: <Unknown>
4
5      PRIOR APPLICATION DATA:
6      APPLICATION NUMBER: 09/453,702
7      FILING DATE: 03-DEC-1999
8      APPLICATION NUMBER: 60/110,955
9      FILING DATE: 04-DEC-1998
10
11     ATTORNEY/AGENT INFORMATION:
12     NAME: Seay, Nicholas J.
13     REGISTRATION NUMBER: 27386
14     REFERENCE/DOCKET NUMBER: 960296.95017
15
16     TELECOMMUNICATION INFORMATION:
17     TELEPHONE: (608) 251-5000
18     TELEFAX: (608) 251-9166
19
20     INFORMATION FOR SEQ ID NO: 102:
21     SEQUENCE CHARACTERISTICS:
22     LENGTH: 7886
23     TYPE: nucleic acid
24     STRANDEDNESS: double
25
26     TOPOLOGY: linear
27
28     MOLECULE TYPE: DNA (genomic)
29
30     SEQUENCE DESCRIPTION: SEQ ID NO: 102:
31
32     US-10-114-170-102
33
34     Alignment Scores:
35     Pred. No.:          4.84e-65          Length: 7886
36     Score:              701.50           Matches: 163
37     Percent Similarity: 60.10%          Conservative: 66
38     Best Local Similarity: 42.78%       Mismatches: 135
39     Query Match:       34.74%           Indels: 17
40     DB:                14
41
42     US-09-889-756A-2 (1-412) x US-10-114-170-102 (1-7886)

```

Alignment Scores:					
Pred. No.:	4,846-65	Length:	7886		
Score:	701.50	Matches:	163		
Percent Similarity:	60.10%	Conservative:	66		
Best local Similarity:	42.78%	Mismatches:	135		
Query Match:	34.74%	Indels:	17		
DB:	14	Gaps:	7		
US-09-889-756A-2 (1-412) x US-10-114-170-102 (1-7886)					
QY	7	LysAlmctArgAlaAlaAlaLeuAlaAlaAlaValAlaLeuValLeuSerSerCysGly	26		
DB	6823	AAATATAAGAAAGTATATATAGCAACATCTGTAGTGGCAATGCTGCTTATCTGGTGT---	6767		
QY	27	LysGlyGlyAspAlaAlaGlnGlyGlyGlnProAlaGlyArgGlyAlaProAlaProVal	46		
DB	6766	-----GATTAATACCCAAAGTAAACAATTCATCCCGCTCAAGAAACAGAA-----	6725		
QY	47	ValGlyValValThrValHisProGlnThrValAlaLeuThrValGlnLeuProGlyArg	66		
DB	6724	GTCGGCGTGTGTACAGTCAAAATCTCAACCGCGTCTGGTAGTACAGATTAATTAACCGGACGC	6665		
QY	67	LeuGlnSerLeuArgThrAlaAspValAlaGlnValGlyGlyTleIleGlnLysArg	86		
DB	6664	ACCACTGCTGCCCTCAGTCCGAACTACGTCGCGAGTGTGGGGAAATTATCCAGAAACGC	6605		
QY	87	LeuPheGlnGlnGlySerTyrValAlaArgAlaGlnGlnProLeuTyrGlnIleAspSerSer	106		
DB	6604	TTATTTAAGAAAGGATCTGCTGTCAAGCTGGACACGCCGCTTACCAAGATTGATCGCGCC	6545		
QY	107	ThrTyrGlnAlaAsnLeuGlnSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeu	126		
DB	6544	AGTTATCAGGCTGCATGGAATGAACCTCGGCGCATTAACAACAGCACAGGCA---CTG	6488		
QY	127	AlaIysAlaAspAlaAspLeuAla---ArgTyrLysProLeuValAlaAlaGlnAlaVal	145		
DB	6487	GTAATAAGCCGATGGCCAGAAAGCCGAGCGTTATGCCGATCTGTAAAGAGAAACGTTGTT	6428		
QY	146	SerArgGlnGlnuTyrAspAlaAlaValThrAlaLysArgSerAlaGlnAlaGlyValLys	165		
DB	6427	TCACACACGATCGTGAATGCTCAAGTCTTACCTTGACCAAGATTAAGCCAGTGTACGC	6366		
QY	166	AlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsnArgSerArgIleThrAla	185		
DB	6367	CGGAATAAAGCCGACCTGGAACTCGCGCATTTATCTTGACTGGACACGCTAACCCCA	6308		
QY	186	ProIleSerGlyPheIleGlyGlnSerLysValSerGlnGlyThrLeuLeuAsnAlaGly	205		

Db	6307	CCGATTTCCAGGCGGCATTTGGCATTTTCGTCGGTAACCCCTGGCAGCATGTCAGCGCG	6248
Qy	206	AspThrThrValLeuAlaThrTrlLeArgInThrAsnProMetTyrValAsnValThGln	225
Db	6247	CAAGATCAGCGGTTTAAGCAGCTATTTCGGTTCGTGATCAAAATGATGTCGACTCCTCAGC	6188
Qy	226	SerAlaSerGluValMetLysLeuArgArgGlnIleAlaGluGlyLysLeuLeuAla	245
Db	6187	TCCAGTGTCCATTTATTACGTCCTCGTAAACAG-----TCACGTGGCAACCAAC	6140
Qy	246	AspGlyValIleAlaValGlyIleLysPheAspAspGlyThrValTyrProGluLysGly	265
Db	6139	AGTGACACCATGAGCGTCCTCACTATTCTCGAAGATGGCACCAACTCAGCGAAAAAGG	6080
Qy	266	ArgLeuLeuPheAlaAspProValValAsnGluSerThrGlyGlnIleThrLeuArgAla	285
Db	6079	CGTGTGAAGCTCACCGAAGTCGCTGTAGATGACTACCGGTCGGACATTCAGGCGC	6020
Qy	286	AlaValProAsnAspGlnAsnIleLeuMetProGlyLeuTyrValArgValLeuMetAsp	305
Db	6019	ATTTTTCCCAATCCACAACAGCATTTATTACCGGGAATGTTTCTTCGCGCTCGTGCAT	5960
Qy	306	GlnValAlaValAspAsnAlaPheValValProGlnGlnAlaValThrArgGlyAlaLys	325
Db	5959	GAAGCGCGTGAAGAAGACGCTATTCTCGCGCCGCAACAGGGCGCTCACGCGGAGCTAAA	5900
Qy	326	Asp-----ThrValMetIleValAsnAlaGlnGlyLysMetGluProArgGluValThr	343
Db	5899	GGCAATGCACATGGCGGTGGTGTAAATTAAGACATTAAGATGAGCAGCGCAACCTCGAA	5840
Qy	344	ValAlaGlnGlnGlnGlyThrAsnTrpIleValThrSerGlyLeuLysAspGlyAspLys	363
Db	5839	ACGGGAGAAATATGTGGATTAATTCGCTGCTTAACGGCTGCACAGCGGAGCACCGA	5780
Qy	364	ValValValGluGlyIleSerIleAlaGlyIleThrGlyAlaLysLysValThrProLys	383
Db	5779	CTGATTGTGAAGGT-----TCGCCAAAGTCACTTCAGGCGCAGAGCGTCAAGCTGTT	5726
Qy	384	Glu 384	
Db	5725	GAA 5723	

```

RESULT 2
US-09-912-020-121
; Sequence 121, Application US/09912020
; Patent No. US20020045592A1
; GENERAL INFORMATION:
; APPLICANT: Zyskind, Judith
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Trawick, John
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Froelich, Jamie M.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; FILE REFERENCE: ELITRA, 001DVI
; CURRENT APPLICATION NUMBER: US/09/912,020
; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: 09/492,709
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/117,405
; PRIOR FILING DATE: 1999-01-27
; NUMBER OF SEQ. ID NOS: 465
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 121
; LENGTH: 1395
; TYPE: DNA
; ORGANISM: E. Coli
US-09-912-020-121
Alignment Scores:

```

Pred. No.: 2,28e-37 Length: 1395
 Score: 435.50 Matches: 124
 Percent Similarity: 47.94% Conservative: 62
 Best Local Similarity: 31.96% Mismatches: 176
 Query Match: 21.57% Indels: 26
 DB: 9 Gaps: 7

US-09-889-756a-2 (1-412) x US-09-912-020-121 (1-1395)

QY 22 LeuSerSerCyglYlyglYglYaspaLaalaglInglyLy-GlnProAlagly----- 39
 DB 240 CTCGCCGAGTGCAGGCCCGGCGCAAAACAGCCGACGAAATCGCCAGCGGTGCTCG 299
 QY 40 -----ArgLualProAlaProValglYValThraInSproGlnTh 56
 DB 300 ACGTGTATGCGTTCGCCGCTTAGCCCGGTTCAGCGCGGACCGCCGTAAGAACAGCC 359
 QY 56 rValAlaLeuThraValgluLeuProglYArgLeuGlnSerLeuArghThraLaaspValAr 76
 DB 360 AGTTCGCGGTACCTCACCGGCTTGACCATTAACCGCTTAATCCGTTACGCTGCG 419
 QY 76 gAlaGlnValglYlyglYleIleIleGlnLyArgLeuPheGlnGlnGlySerTyValArGAl 96
 DB 420 CAGCCGCGTGCAGCGCCCACTGATAGCTTACATTTCCAGGAAGCCAGCAGGTCAAGC 479
 QY 96 agLYGlnProLeuTyrgInIleAspSerSerThTyrgInuLaasnLeuGlnSerAlaAr 116
 DB 480 AGCGATTTACTGGCAGAAATTAAGCCCGACGACGTTCAAATTCATTAGCACAAGCCCA 539
 QY 116 gAlaGlnLeuAlaThraLaaglInaIleThraLeuAlaLysAlaAspLeuAlaArGTY 136
 DB 540 GGGCCAACTGGCAAAAGATTAAGCAGCGCTTCCACGCGCCGCTGACCTGCGCGCTTA 599
 QY 136 rLYeProLeuValAlaLaaglInaValaLseArGlnGlnTyraAspaLaalValThraI 156
 DB 600 TCACCAACTGGCAAAACCAATCTCGTTTCCGCGAGAGCTGGATGCGCAACAGCGGCT 659
 QY 156 aLYsArSerAlaGlnuIlaaglYValLySaLaaglInaLaalIleLysSerAlaGlyI 176
 DB 660 GGTCAGTGAACCGAAGCAGCATTAAGCTGATGAAGCAAGCTTCCAGCGGAGCT 719
 QY 176 eAsnLeuAsnArSerArghIleThraIaProIleSerGlyPheIleGlyGlnSerLySa 196
 DB 720 GCAATCGATGAGCGCGGATTAACCGACAGTGTGCTGCGCTTGTCTCAAGCAGGT 779
 QY 196 lSerGlnGlyThraLeuAsnAlaGlyAspThThraValLeuAlaThraIleArghInTh 216
 DB 780 TGATGTTGGTACCAAAATCTCCAGTGTGATACACCGGATCGTGTGATACCCAGAC 839
 QY 216 rAsnProMetTyraValAsnValThraGlnSerAlaSerGlnValMetLysLeuArGAl 236
 DB 840 GCATCTATGATTTAGCTTTTACCTTCGCGGAAGGATATGCTTACCTGATGCGAGC 899
 QY 236 nIleAlaGlnGlyLysLeuLeuAlaAlaAspGlyValIleAlaValglYle--Lysph 255
 DB 900 GCAGAAAGCGGAAACCGCTGTGTGAAGACCTGGGATCGACCAACTCGAAGAAATT 959
 QY 255 eAspAspGlyThraValTyProGlnuLyGlyArgLeuPheAlaAspProValValAs 275
 DB 960 AAGTGAAGGACG-----CTGTAAAGCTTAAGTAAACCAAAATCGA 998
 QY 275 nGlnSerThraGlyGlnIleThraLeuArgAlaAlaValProAsnAspGlnInleLeu 295
 DB 999 TGGCACATCCGATGATTAAGTAAAGGACGACGCTTAAATATAGATATGCTGCTT 1058
 QY 295 tProGlyLeuTyraVal-----ArgValLeuMetAspGlnValAlaValaAspAlaaph 313
 DB 1059 TCCCAATCAGTTTGTAAACGCGCATGTTAGTGAC-----ACCGAACAAACGCGCT 1112
 QY 313 eValAlaProGlnGlnAlaValThraArgGlyAlaLys--AspThraValMetIleValAs 332
 DB 1113 AGTATCCCAACAGCGCCCTCGCAATATGGCAATGAAGGCCATTTTGTGCGGTGAA 1172

QY 332 nAlaGlnGlyGlyMetGlnProArghIuValThraValaIleGlnGlnGlyThraSnr 352
 DB 1173 TAGCGAAACAAAGGATCGCAAAACATCTGTGACCGCGGATTCAGACAGTCAAGAAAGT 1232
 QY 352 pIleValThraSerGlyLeuLysAspGlyAspLySaValValaGlnGlyle----- 369
 DB 1233 GGTGATCCGTGACAGTATTTCTGCGGCGATCGCGGTGTACAGACGCGCATGATGCTCT 1292
 QY 370 -----SerIleAlaGlyIleThraGlyAlaLysLysValThraProLyGlnTr 385
 DB 1293 GACCGAAGGCGCAAAAGTGAAGTGTGGAAGCCAGAGCCCATCTCTCGAAGAGAA 1352
 QY 385 pAlaSerSerGlnuSnglnAla 392
 DB 1353 AGCCACGACCGCGAATATACCG 1374

RESULT 3
 US-09-975-719-148
 ; Sequence 148, Application US/09975719
 ; Publication No. US2003022349A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Ausubel, Frederick M.
 ; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
 ; FILE REFERENCE: 00786/361003
 ; CURRENT APPLICATION NUMBER: US/09/975, 719
 ; PRIOR FILING DATE: 2001-10-10
 ; PRIOR APPLICATION NUMBER: US 09/199, 637
 ; PRIOR FILING DATE: 1998-11-25
 ; PRIOR APPLICATION NUMBER: US 60/066, 517
 ; NUMBER OF SEQ. ID NOS: 437
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ. ID NO 148
 ; LENGTH: 1008
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 ; US-09-975-719-148

Alignment Scores:
 Pred. No.: 1.51e-30 Length: 1008
 Score: 370.50 Matches: 80
 Percent Similarity: 58.29% Conservative: 43
 Best Local Similarity: 37.91% Mismatches: 73
 Query Match: 18.35% Indels: 15
 DB: 11 Gaps: 5

US-09-889-756a-2 (1-412) x US-09-975-719-148 (1-1008)

QY 205 GLYAspThThraValLeuAlaThraIleArghInThraSnrProMetTyraValaSnValThr 224
 DB 1 GGCAGGCAAAACCGATGGCGACCGTGCACACACTGACCCGATCTTACGTGACGTACCC 60
 QY 225 GlnSerAlaSerGlnuValMetLysLeuArgArghInIleAlaGlnGlyLysLeu--Leu 243
 DB 61 CACCGCTTCAACCGCCCTGTTCGCGATCGCGCGAACTGCGCAGCGCGCATTTGAGCGC 120
 QY 244 AlaAlaAspGlyValIleAlaValglYleLysPheAspAspGlyThraValTyProGln 263
 DB 121 GCCGCGCAACAGCTCGAAGGTCCTCGAAGCTGGAAGCGGTAGCCAAATACCGCTG 180
 QY 264 LySGlyArgLeuLeuPheAlaAspProValValaSnGlnSerThraGlyGlnIleThra 283
 DB 181 GAAGGCGCTCGAATTCCTCGAGGTTTCGTCGAGGAAGGACCGGCTCGTGCACCATC 240
 QY 284 ArgAlaAlaValaProAsnAspGlnInleLeuMetProGlyLeuTyraValaArgValaLeu 303
 DB 241 CCGCGCGTGTTCGCCAACCCGAAACAGAGCTGCTGCGCGCATGTCTTCACGCGCAG 300
 QY 304 MetAspGlnValaValaValaAspAlaPheValaProGlnGlnAlaValaThraArgGly 323
 DB 301 TTGCAGAGAGCGGTCAAGCAGAGGCGCATCTGCTCCGACAGAGGCGTGAACCGCGAC 360

```
NAME/KEY: misc feature
LOCATION: (40808)..(40810)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (44416)..(44416)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (44975)..(44975)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (45732)..(45732)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (45593)..(45593)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51334)..(51334)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (55369)..(55369)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (100091)..(100091)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
```

```
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (147197)..(147197)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (152500)..(152500)
```

```
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (152530)..(152530)

Alignment Scores:
Pred. No.: 5.98e-17 Length: 1830121
Score: 286.00 Matches: 105
Percent Similarity: 41.01% Conservative: 66
Best Local Similarity: 25.18% Mismatches: 148
Query Match: 14.17% Indels: 98
DB: 14 Gaps: 14

US-09-889-756a-2 (1-412) x US-10-329-960-1 (1-1830121)

Qy 6 PheylsalmetarglaalaalaaleualaalaalvalaaleuvalleuSerSerCys 25
Db 946822 TTTAATATGATATAAAGCGTAATGATATAGCCGAGCCATTGCA----- 946863

Qy 26 GlyLysGlyGlyaspalaalaglInGlyGlyProbaGlyArgGluAlaProAlaPro 45
Db 946864 -----GGAATGCCA-----GAAATCTTCAAGCCCA 946887

Qy 46 ValValGlyValValThrValHsProGlnThrValAlaLeuThrValGluLeuProGly 65
Db 946888 GTAACC---GCACCTGAAGTTCAACCGCGTAATGACGCGCAGTTATTACACAAACAGT 946944

Qy 66 ArgLeuGluSerLeuArgThrAlaAspValArgAlaGlnValGlyGlyLeuGlnLys 85
Db 946945 CTGTGGTCCTCAATCAAGCGCAATGCTCAGTACACAAATGCGGCGGCTTCCAA 947004

Qy 86 ArgLeuPheGlnGluGlySerTyrrValArgAlaGlyGlnProLeuTyrrGlnIleAspSer 105
Db 947005 GTACTTGTCAAAATGACCAAAATGTGCAAAAGGTGAGTGCTTGTGGAGCTTGATGT 947064

Qy 106 SerThrTyrrGluIleAsnLeuGluSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThr 125
Db 947065 TCTGTGAACGACCTATCTACAGCTGCTCAGGACCAATTATCAGACCTTCGCAAACT 947124

Qy 126 LeuAlaLysAlaAspAlaAspLeuAlaArgTyrrLysProLeuValAlaAlaGluAlaVal 145
Db 947125 -----TACCAACGTTATGTGGTTATTAAATGCAATGCTGTA 947163

Qy 146 SerArgGlnGluTyrrAspAlaAlaValThrAlaLysArgSerAlaGluAlaGlyValLys 165
Db 947164 TCACGTCACAAATGTGATTAACGCAAAAGCGCTTATGATGCTCAGTACGTAAGTAA 947223

Qy 166 AlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsnArgSerArgIleThrAla 185
Db 947224 TCTCTAAAGCGACCAATTGAA-----CGTGGTAAATTTGTTGCG 947262

Qy 186 ProIleSerGlyPheIleGlyGlnSerLysValSerGlnGlyThrLeuLeuAsnAlaGly 205
Db 947263 CCATTGTGATGCGCAACAGGATATGTGAAATCAAGTTGACAAATATGAAATGTGGA 947322

Qy 206 AspThrThrValLeuAlaThrIleArgGlnThrAsnProMetCyrValAsnValThrGln 225
Db 947323 -----ACGAAATGTGCTGCTGAGAAAGTACTCAATGAAAGTGGATTGCTCTT 947376

Qy 226 SerAlaSerGluValMetLysLeuArgGlnIleAlaGluGlyLysLeuLeuAlaAla 245
Db 947377 TCACAAATGATTATTAATAATTACAT-----ATCGGTACAGCGCTTACAGCG 947424

Qy 246 AspGlyValIleAlaValGlyIleLysPheAspAspGlyThrValTyrrProGluLysGly 265
Db 947425 ACAACAGATGCTCGCTGGCGGAACATTT-----TCAGCT 947460

Qy 266 ArgLeuLeuPheAlaAspProValValAsnGluSerThrGlyGlnIleThrLeuArgAla 285
Db 947461 CGAATACGCGCATTCGCTGCAATTAATATCAACAGGTTTACTGATGTTCAAGGCT 947520

Qy 286 AlaVal---ProAsnAspGlnAsnIleLeuMetProGlyLeuTyrrValArgValleuMet 304
Db 947520
```

```

Db 947521 ACTTTGATCCCTGAAGATGGCATTAATTCCTTCAGGTATGCTCTCGCTTACGCATT 947580
Qy 305 AsglnValAlaValAspAsnAlaPheValValProGlnGlnAlaValThr----- 321
Db 947581 GCACTTCCAACTGAAACAAATCAAGTTGTCGTCACAGTAGCTATTAGTACAAATATG 947640
Qy 321 ----- 321
Db 947641 TATGGCGAATTGGCTATTATTGAACTTATCTGAAGAAAGAAAGAAATATGCA 947700
Qy 322 -----ArgValAlaValAspThrValMetIleValAsn 332
Db 947701 GGTAATGAAATAATGATGCTCTCTATGTCGAAACAGATCCAGTATTATTAATAAGAT 947760
Qy 333 AlaGlnGlyMetGluProArgGluValThrValAlaGlnGlnGlyThrAsnTrp 352
Db 947761 CGTCAAGGT-----GTTATGCTCAATTACAGGAAATGAA--- 947796
Qy 353 IleValThrSerGlyLeuLysAspGlyAspLysValValGluGlyIleSerIleAla 372
Db 947797 -----GTTAAAGTGGAGATAAATATTATACAGCGGCTCAGCAA----- 947835
Qy 373 GlyIleThrGlyAlaLysLysValThrProLysGluTrpAlaSerGlu 389
Db 947836 GGATTTGGTAATGGAAGTCTTGTC-----GAATGATTTAAAAAGAC 947877

RESULT 5
US-09-975-719-171/c
; Sequence 171, Application US/09975719
; Publication No. US20030022349A1
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Rahme, Laurence G.
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; FILE REFERENCE: 00786/361003
; CURRENT APPLICATION NUMBER: US/09/975,719
; CURRENT FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 09/199,637
; PRIOR FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: US 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 171
; LENGTH: 1512
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-975-719-171

Alignment Scores:
Pred. No.: 6,65e-12 Length: 1512
Score: 200.00 Matches: 47
Percent Similarity: 65.93% Conservative: 13
Best Local Similarity: 51.65% Mismatches: 25
Query Match: 9.91% Indels: 6
DB: 11 Gaps: 3

US-09-889-756a-2 (1-412) x US-09-975-719-171 (1-1512)
Qy 14 LeuAlaAlaValAlaValAlaLeuValLeuSerSerCysGlyLysGlyValAspAlaAlaGln 33
Db 255 CTGGGGCGCCCTGGCTGCTTCTCTGCTGGGCTC-----GAAGAACAGCGGAC 205
Qy 34 GlyLysGlnProAlaGlyArgGluAlaProAlaProAlaValGlyValValThrValHis 53
Db 204 ACCGGGAGAGCTGCG-----GAGGCCCGCGCGAG-----GTCCGGGTGATGCTGCCAGG 154
Qy 54 ProGlnThrValAlaLeuThrValGluLeuProGlyArgGluGluSerLeuArgThrAla 73
Db 153 CCGGGCCCATCGCATCCAGCGAGCTGCGCGAGCCCTGGAGAGCTATCCAGCGCC 94
Qy 74 AspValArgAlaGlnValGlyIleIleGlnLysArgLeuPheGlnGluGlySerTyr 93

```

```

Db 93 GAGGCGCGCGCGCGCGCTGCGCGCATGCTACCCGTCGCTGTACGAGGAAGCGCAGCA 34
Qy 94 ValArgAlaGlyGlnProLeuTyrGlnIleAsp 104
Db 33 GTCGCGCGCGCACCGTCTGTTCCAAATCGAC 1

RESULT 6
US-09-738-626-3315
; Sequence 3315, Application US/099738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, MAKIO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 3315
; LENGTH: 1977
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
; US-09-738-626-3315

Alignment Scores:
Pred. No.: 4.8e-09 Length: 1977
Score: 175.00 Matches: 102
Percent Similarity: 39.70% Conservative: 81
Best Local Similarity: 22.13% Mismatches: 194
Query Match: 8.67% Indels: 84
DB: 10 Gaps: 16

US-09-889-756a-2 (1-412) x US-09-738-626-3315 (1-1977)
Qy 5 AlaPheLysAlaMetArgAlaAlaAlaLeuAlaAlaValAlaValLeuValLeuSerSer 24
Db 667 GCTCTCAAGGACGCTGCTGAGCAAGCAGATGCTCTGAATTTGAACAGATCGCGCGGAC 726
Qy 25 CysGlyLysGlyLysAspAlaAlaGlnGlyGlnProAlaGlyArgGluAlaProAla 44
Db 727 TTGGAGCTGCTCAACAC-----GATCCAGCAACCTCAAC 762
Qy 45 ProValValGlyValValThrValHisProGlnThrValAlaLeuThrValGluLeu--- 63
Db 763 GATGCTCATTTGCTGCTGCTGATGAGAGAAATCTTGGCTTCCGGAATCCGAATCGACTGCT 822
Qy 64 -----ProGlyArgLeuGluSerLeuArgThrAlaAspValArgAlaGlnVal 79
Db 823 CAAGCCCGCGCTGAGGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 876
Qy 80 GlyIleIleIleGlnLysArgLeuPheGlnGluGlySerTyrValArgAlaGlyGlnPro 99
Db 877 GCAAGCCTGGAACAATATGCTTCCAAACCTCCACAGTGGCTTACCAAGATCAGACT 936
Qy 100 LeuTyrGlnIleAspSerSerThrTyrGlnAlaAsnLeuGluSerAlaArgAlaGlnLeu 119
Db 937 TACTTGAG-----TCTTACACCGCTTTGAGAGAACTGAGCGAGCGGCTGCA 984

```



```

Qy 120 AlaThrAlaGlnAla-----ThrLeuAlaLysAla 129
Db 985 TCACACCTAGAGCTCTAGAAATAGCCGAGGATCTACATTGATTCTCGGAAAAGTT 1044
Qy 130 AspAlaAspLeuAlaArgTyrLysProLeuValAlaAlaGluAlaValSerArgGln 149
Db 1045 GACTCAGAACTAGCCCGCCGACGACGCCCTT---GCCGAAGCCCACTGACACACAA 1101
Qy 150 TyrAspAlaAlaVal-----ThrAlaLysArgSerAlaGluAlaGlyValLysAla 166
Db 1102 ---GACGACGACCTTGGCTCGAAGCCGCGACGCTTCCACCAACCAATGGAAGCC 1158
Qy 167 AlaGlnAlaAlaLysSerAla----- 174
Db 1159 CAATCAGCCGCTCATGATGACGCTTTAGTTGGATCAGTAAATGAAGCCGCCACC 1218
Qy 175 -----GlyTLeuLysLysArgSerAlaGlyThrLeuLys 188
Db 1219 AGATCCACGTCCTCAGCTGCGATGATCAACACACCCGTCGCTCCCATCTCA 1278
Qy 189 GlyPheLe-----GlyLysSerLysValSerGlyThrLeuLys 203
Db 1279 GGCATTGTTATCTCCGACGACGACCCAGGTCAACGACGCGGCGCATCTTGAAT 1338
Qy 204 AlaGlyAspThrThr-----ValLeuAlaThrIleArgIleThrAspMetTyr 220
Db 1339 GTTCTGATGATTCGCAACTGAAGATCACCGCGAATGAAAGAGCGGAGATCAGAAC 1398
Qy 221 ValAsnVal-----ThrIleSerAlaSerGluValMetLysLeuArg 234
Db 1399 GTCAACATCGGATCCCGGTACCTTCACTACCCCATCGGACCAAAAGATTCGCC 1458
Qy 235 ArgIleIleAlaGluLysLeuLeuAlaAlaAspGlyValIleAlaValGlyIleLys 254
Db 1459 GCGCGAGATCAAAAGTCTCCCATTCGACGCT----- 1491
Qy 255 PheAspAspGlyThrValTyrProGluLysGlyArgLeuPheAlaAspProValVal 274
Db 1492 -----GCCGCAATGCCCGCCAGCTACAGGTAAAGAGTGGCCGACGCCACAAAC 1542
Qy 275 AsnGluSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeu 294
Db 1543 ACCAACTACGATCCTTCCCATGAAATTTCCGTACCGGCGACCGCAAGGCGCTC 1602
Qy 295 MetProGlyLeuTyrValArgValLeuMetAspGlnValAlaValAspAsnAlaPheVal 314
Db 1603 AACCTGCGGAGATCCGCTCGAGTACGATCGTCCATGAATCGACACACGATCTGACC 1662
Qy 315 ValProGlnGlnAlaValThrArgGly-----AlaLysAspThrValMetIleValAsn 332
Db 1663 GTTCTCTGGAAGTGTGTACAAAATGATGACGCGCAAAAGCGTGTTCATCATCAGC 1722
Qy 333 AlaGlnIleGlyMetGluProArgGluValThrValAlaGlnGlnGlyThrAspTyr 352
Db 1723 GACGACAACAAAGTAAAGTAAAGTAAACACCTGATCCATGCTGATGATC 1782
Qy 353 IleValThr---SerGlyLeuLysAspGlyAspLysValValGluGlyIleSerIle 371
Db 1783 GCAATCAGCGGTGTGAAATTTCAAGAACGCTGATGCTCACCCGCTGGAATATC 1842
Qy 372 AlaGlyIleThrGlyAlaLysValThrProLysGlnThrPheLysSerGluAsnGln 391
Db 1843 CGGGGCTCATCGAAGAA-----ACTGTGAACTTCACGCAAGTACGGTGGACGAG 1893
Qy 392 AlaAlaAlaProGlnSerGlyValGlnThrAlaSerGlnAlaLysThrAlaSerGluAla 411
Db 1894 GCGGCGGCTCTTTAGTCTCGGCGCCCTTTGACCTCGACGCGCTTCGCTTTCGCC 1953
Qy 412 Glu 412
Db 1954 AAG 1956

```

```

RESULT 7
US-09-738-626-1/c
; Sequence 1, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1
; LENGTH: 3309400
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

```

```

Alignment Scores:
Pred. No.: 0.000129 Length: 3309400
Score: 175.00 Matches: 102
Percent Similarity: 39.70% Conservative: 81
Best Local Similarity: 22.13% Mismatches: 194
Query Match: 8.67% Indels: 84
DB: Gaps: 16

```

```

US-09-889-756a-2 (1-412) x US-09-738-626-1 (1-3309400)
Qy 5 AlaPheLysAlaMetArgAlaAlaAlaAlaAlaValAlaLeuValLeuSerSer 24
Db 3196520 GCTCTCAAGGACCTGCTGACGACGATGCTGCTGAATTAAGACTGCCGCGAC 3196461
Qy 25 CysGlyLysGlyLysAspAlaAlaGlnGlyGlnProAlaGlyArgGluAlaProAla 44
Db 3196460 TTGGGACTGCTCAACAC-----GATGGACGCAACCTCAAC 3196425
Qy 45 ProValValGlyValValThrValHisProGlnThrValAlaLeuThrValGluLeu--- 63
Db 3196424 GATGTCATTGCTGCTGCTCATGAGAGAGAAATCTTGCTTCGCGAATCCGAATC 3196365
Qy 64 -----ProGlyArgLeuGlnSerLeuArgThrAlaAspValArgAlaGlnVal 79
Db 3196364 CAAGCCCGGCTGACGATGACCTGAGGCGATGCTGCGACGTA-----GCAAGGTT 3196311
Qy 80 GlyGlyIleIleGlnLysArgLeuPheGlnGlnGlySerTyrValArgAlaGlnPro 99
Db 3196310 GCAAGGCTGGAACATCATCTGCTCCAAACCTCAGCGCTGACGTAAGTCACT 3196251
Qy 100 LeuTyrGlnIleAspSerSerThrTyrGluAlaAsnLeuGluSerAlaArgAlaGlnLeu 119
Db 3196250 TACTTCGAC-----TCCTACACCGCTTTGAGAGAACTGAGCGACGCGTCA 3196203
Qy 120 AlaThrAlaGlnAla-----ThrLeuAlaLysAla 129
Db 3196202 TCACACCTGAGACTTGAATAGCCGAGGATTAATGATGATTCACCTCGGAAAAGTT 3196143
Qy 130 AspAlaAspLeuAlaArgTyrLysProLeuValAlaAlaGluAlaValSerArgGln 149

```

```

Db      3196142  AACTCAGAACTAGACCGCGCACAGCGGCCGCTT---GCCGAAGCCACTCAGCAACAACAA 3196086
QY      150  TyrhpaalaaVala-----ThralaalsarSerlaagluaglyVallysAla 166
      |||||
Db      3196085  ---GACGAGACACTTGGCGCTCGAGACC CGCAGACTTTCACCCAAACAAATTGGAAGCC 3196029
QY      167  AlaaglnlaaIleIleYsSerAla----- 174
      |||||
Db      3196028  CAATCAGAGCGCATGATGACAGCTTTAGGTTGGCATCATGATATGAAGCCGCCACC 3195969
QY      175  -----GlyIleasnleuasnArSerArgIleThraAlaProIleSer 188
      |||||
Db      3195968  AGATCCACAGTCCACAGCTCGGATGGATGTATCAACAACACACCGTTCCGCTCCCATCTCA 3195909
QY      189  Glyhelele-----GlyInSerIysValSerGluGlyThrLeuIleuasn 203
      |||||
Db      3195908  GGCATTGTCTTCATCCGTGACGAGCACGCCAAGGTCAACACAGACGCGCGCACTGTGGAGT 3195849
QY      204  AlaGlyAspThrThr-----ValleuAlaThrIleArgGlnThraAspProMetCys 220
      |||||
Db      3195848  GTTGCTGATGATTCCGAACTGAAGATCAACCGGAATGTAAAGAACGCGAGATCAGCAAC 3195789
QY      221  ValaenVal-----ThrGlnSerAlaSerGluValMetIysLeuArg 234
      |||||
Db      3195788  GTCAACCATCGGATCCCGCGCTCACTCTCACTACCCCATGACCGGAAACCAAGAAATTGCC 3195729
QY      235  ArgGlnIleAlaGluGlyLysLeuIleuAlaAlaAspGlyValIleAlaValGlyIleLys 254
      |||||
Db      3195728  GGCCACGATATCCAAAGTCTCCCGCCCTTCGACCT----- 3195696
QY      255  PheaAspArgIlyThrValIlyrProGluIysGlyrLeuLeuPheAlaAspProValVal 274
      |||||
Db      3195695  -----GCCGCCAGTGGCCCGACGTAACAGGTGAAGAGCTGCCGCGACGCCCAAC 3195645
QY      275  AsnGlnSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeu 294
      |||||
Db      3195644  ACCAACACTGACAGTCACCTCCCATCGAAATTTCCGTACCCGGGACCGGGAAGGCTTC 3195585
QY      295  MetProGlyLeuIlyrValAlaArgValIleuMetAspGlnValAlaValAspAsnAlaPheVal 314
      |||||
Db      3195584  AACCTCGGCGGATCGCTCGAGTACGATCGTCATGAATCGCACACGACTGAC 3195525
QY      315  ValProGlnGlnAlaValThraArgGly-----AlaIysAspThrValMetIleValAsn 332
      |||||
Db      3195524  GTTCTTTGGAAAGCTGTACAAAATGATGACGGCAAGACCGCTGTTTATCATCGC 3195465
QY      333  AlaGlnGlyIleMetGluProArgIleValThraValAlaGlnGlnGlnGlyThraAspTirp 352
      |||||
Db      3195464  GACGCAACAAGTGAAGAAAGTGAAGTAAACACGCTGAATCCGATGACTTTGATATC 3195405
QY      353  IleValThr-----SerGlyLeuIysAspGlyAspIysValAlaValGluGlyIleSerIle 371
      |||||
Db      3195404  GCAGTCACGCGGTGCGGAATTCAAGAAACGCTGAGTGCACACCGCTCGGAACATC 3195345
QY      372  AlaGlyIleThrGlyAlaIleLysValIleThrProIysGluTTPAlaSerSerGluAsnGln 391
      |||||
Db      3195344  CGGGGCCCTCATCGAGAA-----ACTGTGAACCTTCAACGAGATACGCTGAAGAG 3195294
QY      392  AlaAlaAlaProGlnSerGlyValGlnThrAlaSerGluAlaIysThrAlaSerGluAla 411
      |||||
Db      3195293  GCGGCGGCTCTTTTAGTGTCTCGCGGCCCTTTTGAACCTCGACGCCCTTCGCTTCTGCC 3195244
QY      412  Glu 412
      |||
Db      3195233  AAG 3195231

```

```

APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA_011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 9752
LENGTH: 930.
TYPE: DNA
ORGANISM: Salmonella typhi
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(930)
NAME/KEY: misc_feature
LOCATION: (1)..(930)
OTHER INFORMATION: n = A,T,C or G
US-09-815-242-9752

Alignment Scores:
Pred. No.: 2,47e-09 Length: 930
Score: 173.50 Matches: 70
Percent Similarity: 42.11% Conservative: 50
Best Local Similarity: 24.56% Mismatches: 117
Query Match: 8.59% Indels: 48
DB: 9 Gaps: 8

US-09-889-756A-2 (1-412) x US-09-815-242-9752 (1-930)
QY 48 G1yValValThrValHisProGlnThrValAlaLeuThrValGluLeu-ProGlyArgLe 67
DB 129 GGGCGGGTGGTGGTTATGGAATTACTATATGCAGTACCGTGCGAGCGGGCGATTAAT 188
QY 67 uGuserLeuArgThrAlaAspValAlaArgAlaGlnValGlyylelleleGlnyArgLe 87
DB 189 CCGCGGG---GAACAGCTCAGCGTAACCCACAGGTTCTGGCGACATTAACCAACTCAA 245
QY 87 upheGlnGluGlySerTyrValAlaArgAlaGlyGlnProLeuTyrGlnIleAspSerSerTh 107
DB 246 TATCAAAAGATATACAGTACTGTTCAACGCTGGGAGCGTTCTCTTGTGCTATTATAAACGCC 305
QY 107 rTyrGlnAlaAsnLeuGluSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAl 127
DB 306 CTTCGATATTGCGAAGCTTAACGCCGACAGTGGCCAAACGCCAGTCCGATCTGGC 365
QY 127 alyVala-----AspAlaAspLeuAlaArgTyrIlyProLeuValAlaAlaGluAlaVa 145
DB 366 AAAAGCGAATTAATGAAGCGGATCGACGTGCAT-----TTATCGGCTAACTAAT 416
QY 145 llsarGgngInuTyrAspAlaAlaValThrAlaIysArgSerAlaGlnIleGlyVally 155
DB 417 CTCTGCTGAAGATCTGGATAGCGCAATCTGAATGTTAAACCATGACGCGCCAGCTGAGA 476

```

```

QY 165 SALAAlaGlnAlaAlaIlelySerAlaGlyIleAsnLeuAsnArgSerAlaIleThrAl 185
Db 477 TGTGGCACTGGCAACGTTAAACAGCGCAGTGGCAGCTTAGCCGACGAAAGTAAAC 536
QY 185 AProIleSerGlyPheIleGlyGlnSerIleValSerIleGlyThrIleLeuAsnAlaG 205
Db 537 GCCGGTATCCGGCTGGGAGCACCACCTTCCACCGCGTACCGTATTAATGCTAGCACC 596
QY 205 YASPThrValIleuAlaThrIleArgIleThrAsnPrometyrValaIleuValThrG 225
Db 597 T-----AAGCCACTTTTCGCCCTGGTATGATAG 623
QY 225 nSerAlaSerGluValMet-----LysLeuArgArg----- 235
Db 624 CCATTCATTATTGATGATATTTGAGAGAAACCAACTACGACCATATCCCGCAAG 683
QY 236 -----GlnIleAlaGlnGlyIleLeuLeuAlaAlaAspGlyValIleAl 250
Db 684 TGAGCCCTGATTAAATACACTGTATAGTGCATATTAAGTTACAGAGGTCAAGTTGCGAG 743
QY 250 aValGlyIleIysPheAspAspGlyThrValIleThrProGluIysGlyArgLeuPheAl 270
Db 744 TATCGGGGGCGCTATTATACGACCAAGCGTGAAGTACCTCCGTTTAGTG----- 795
QY 270 aASPProValaIleAsnGluSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAs 290
Db 796 -----CCCGATATCAACCAACCAAGTACCGTGGTATGCTGGCGCAACGCGTACCG 846
QY 290 pGlnAsnIleLeuMetProGlyLeuIleuIleuIleuIleuIleuIleuIleuIleu 310
Db 847 -----GlnGlnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 878
QY 310 pAsnAlaPheVal 314
Db 879 TATCAGCGCTGCTC 891

```

RESULT 9

US-09-815-242-6071

Sequence 6071, Application US/09815242

Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haeselbeck, Robert

APPLICANT: Ohlsen, Kari L.

APPLICANT: Zyckind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

TITLE OF INVENTION: Identification of Essential Genes in

FILE REFERENCE: ELITRA 011A

CURRENT APPLICATION NUMBER: US/09/815,242

CURRENT FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 6071

LENGTH: 900

TYPE: DNA

```

; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1) ... (900)
US-09-815-242-6071

```

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	7.23e-09	900	71	43	105	60
Percent Similarity:	169.00					
Best Local Similarity:	40.86%					
Query Match:	25.45%					
DB:	8.37%					

US-09-889-756a-2 (1-412) x US-09-815-242-6071 (1-900)

```

QY 52 ValHisProGlnThrValAlaLeuThrValGluLeuProGlyArgLeuGluSerLeuArg 71
Db 160 ATACGGCGCAGAACAGGTTTCTATCACCACAGGTGCGGAGCGATATTTGAGCTGAA 219
QY 72 ThrAlaAspValArgAlaGlnValGlyGlyIleIleGlnIysArgLeuPheGlnGly 91
Db 220 ATAAAGAC-----AAC 231
QY 92 SerTyValArgAlaGlyGlnProLeuTyGlnIleAspSerSerThrTyGlnAlaAsn 111
Db 232 CAGCTGGTTAACCGGGGATCTTCTGCTCAGCATCGATTAACGCGCTTTCAGATGCC 291
QY 112 LeuGluSerIleArgAlaGlnIleuAlaThrAlaGlnAlaThrIleuAlaIleAlaAspAla 131
Db 292 GAGCTGACGACAGCGCTCAGTGGCGAAGCCCAATCTGACCTCGCAAAAGCCAAATAC 351
QY 132 AspLeuAlaArgTyProLeuValAlaAlaGlnAlaIleuAlaIleuAlaIleuAlaIleu 151
Db 352 GAAGCGAATGCTGCTGCTATCTC---TACAAATTTTATTTCTGCCGAGAACTCGAT 408
QY 152 AlaAlaValThrAlaIysArgSerAlaGlnAlaGlyValIysAlaAlaGlnAlaAlaIle 171
Db 409 ACCGTTAACCCTCAATGTTAAAGCAGTACAGGCGCAGTGTGATCCGACAGGAGCGCTG 468
QY 172 LysSerAlaGlyIleAsnLeuAsnArgSerArgIleThrAlaProIleSerGlyPheIle 191
Db 469 AAACAGGCAACATGCGCATGCGGCAACAGGAATTCGCCGCCGCTGATGATGGCTG 528
QY 192 GlyGlnSerIleValSerGlyThrIleuLeuAsnAlaGlyAspThrThrValIleuAla 211
Db 529 ACTAACCTTACTACCGCATCGCGATTCGCTGATACCGG----- 570
QY 212 ThrIleArgGlnThrAsnPrometyrValaIleuValThrGlnSerAlaSerGluValMet 231
Db 571 -----AAACCGCTTTGGCTTGTGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 615
QY 232 -----LysLeuArgArgGlnIleAlaGlnIleIys----- 241
Db 616 GGTATTATTGAAGAAACCAAGTTG---CCCATATCCGGAAGTGCACACAGCAAAATT 672
QY 242 LeuLeuAlaAlaAspGly-----ValIleAlaValGlyIleIysPhe 255
Db 673 ACCCTTATAGTATTAACAAACAGTTACAGGCTACGTTTCCAGTATCGGTCGGCGGATT 732
QY 256 AspAspGlyThrValTyProGluIysGlyArgLeuLeuPheAlaAspProValaIleAsn 275
Db 733 TATGATCAGAGCGTTTAAAGTACATCAACCTTAT-----CCGATGTTAAA 780
QY 276 GluSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAspGlnAsnIleuMet 295
Db 781 CCTAACGTCCTCGGTTGAGCTCGCCCAACCGTTCC----- 819
QY 296 ProGlyLeuTyValArgValLeuMetAspGlnAlaAlaValaIleAspAsnAlaPheVal 314
Db 820 -----GTCCGCTTGGCGCTGGATTAAGTCCAGCGATGTCACTGTTGGTG 864

```

RESULT 10

```

US-09-815-242-7815
; Sequence 7815, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyckind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7815
; LENGTH: 858
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(858)
; US-09-815-242-7815

Alignment Scores:
Pred. No.: 3,4e-08 Length: 858
Score: 162.50 Matches: 75
Percent Similarity: 40.13% Conservative: 45
Best Local Similarity: 25.08% Mismatches: 127
Query Match: 8.05% Indels: 52
Gaps: 10

US-09-889-756a-2 (1-412) x US-09-815-242-7815 (1-858)
QY 31 AAlaAagInGlyGInProAlaGlyArgGluAlaProAlaValGlyValAl 50
Db 12 GCTTCGCTCGTCGTCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 71
QY 51 ThrValHisProGInThValAlaLeuThrValGluLeuProGly-ArgLeuGInSe 70
Db 72 CTGGCGCTACTACATGCTCTGCGCTGACCCCGCGCGATGCCGGTGGCGGAC 126
QY 70 uArgThrAlaAspValArgAlaGInValGlyIleIleGInuysArgLeuPheGInG 90
Db 127 -----GTGCTGTGTGTGGCGCGCGAGCTGTCCGCTGGGTGACCGACTGGAGTCAAGGA 182
QY 90 uGlySerTyValArgAlaGInProLeuTyGInIleAspSerSerThTyGInAl 110
Db 183 CAACAGGTGTCAAGTGTGGCGAGCTGTGATGCGCATCGACGAAACGCTACAGGC 242
QY 110 aAsnLeuGInSerAlaArgAlaGInLeuAlaThrAlaGInAlaThrLeuAlaLysAlaAs 130
Db 243 CAACCTGAGCAGGCGCGCGCGCTGCGCGAGACCGCTGACCAAGCAATACCTGTTGGCGCA 302
QY 130 pAlaAspLeuAlaArgTyLysProLeu-----ValAlaGInAlaValSe 146

```

```

Db 303 GAACGAGCGCGCGCGAGCGCGCTGCGCATGCGCGCATGACGCGGAGGACAGCA 362
QY 146 rArgInGInTyTrAspAlaAlaValThrAlaLysArgSerAlaGInAlaGlyValLysAl 166
Db 363 AAACGCCCAATCAATCCGCCATC-----GCTCGCGCAGATACAGCA 407
QY 166 aAlaGInAlaAlaIleLysSerAlaGlyIleAsnLeuAsnArgSerArgIleThrAlaPr 186
Db 408 GCGCTGTGGCCAGTGAAGATCCCGCACTCAACCTCAACGCGAGCGAGCGCGCGC 467
QY 186 oLleSerGlyPheIleGInSerLysValSerGlyIleThrLeuAsnAlaGlyAs 206
Db 468 GCGCAACGCGCAGGTATCAACCACTCGCGCTGCGCGCAGGCACTACCGCACGCGGCGCA 527
QY 206 pThrThrValLeuAlaThrIleArgInThraN-----ProMetTyValaAsnVa 223
Db 528 G---GCGGTATGTGCTGTGTCACGACAGTCTCTATGTGTGCTCCCTACTTCGAGGA 584
QY 223 lThrGInSerAlaSerGluValMetLysLeuArgArgGInIleAlaGInuGlyLysLeu 243
Db 585 AACCAAGCTCGCGGATCCGCGCTGCGCATGCGCGCGCGAGTCT-----CGCCTGAT 635
QY 243 uAlaAlaAspGlyValIleAlaValGlyIleLysPheAspAspGlyThrValTyProG 263
Db 636 GAGCGGCGATCAGCGCATC-----GACGCGCAGC----- 663
QY 263 uLysGlyArgLeuLeuPheAlaAspProValValaAsnGInSerThrGlyInIleThrLe 283
Db 664 -----GTGGAGACATACGACGCGGTATCACCGA 692
QY 283 uArgAlaAlaValProAsnAspGIn-----AsnIle----- 293
Db 693 CCGCAACTCGACCGCGGACCGGACGAGTGTGCGCAATGTGACGCCACCTTCACTGGGT 752
QY 294 ----LeuMetProGlyLeuTyValArgValLeuMetAspGInValAlaValAsp 310
Db 753 GCGCCTGCGCGAGCGGTATTCGCGTGCATCCGCTGACACGAGTGTGCGCGGAT 807

RESULT 11
US-10-329-960-1/c
; Sequence 1, Application US/10329960
; Publication No. US20030099277A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Fragment
; FILE REFERENCE: PB186P1
; CURRENT APPLICATION NUMBER: US/10/329,960
; PRIOR FILING DATE: 2003-01-02
; PRIOR APPLICATION NUMBER: US 09/643,990
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 08/487,429
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a, t, g or c
; NAME/KEY: misc_feature
; LOCATION: (10150)..(10150)

```

```

OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (129298)..(29298)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36543)..(36543)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36551)..(36551)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36636)..(36636)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (40808)..(40810)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44416)..(44416)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44905)..(44905)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44975)..(44975)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45593)..(45593)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45732)..(45732)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51334)..(51334)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (55369)..(55369)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a, t, g or c

```

```

FEATURE:
NAME/KEY: misc_feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (100091)..(100091)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:

```

Db 954023 ATGGCAGTGGTATCTAATGAACAATGTGGCTAGAGCTAATTCAAAGAAACCCAATTA 953

NUMB OF SEQ TD NOS: 14110
PRIOR FILING DATE: 2001-02-1

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

```

SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 7049
LENGTH: 1173
TYPE: DNA
ORGANISM: Haemophilus influenzae
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(1173)
US-09-815-242-7049

Alignment Scores:
Pred. No.: 2,32e-07 Length: 1173
Score: 156.50 Matches: 71
Percent Similarity: 37.20% Conservative: 51
Best Local Similarity: 21.65% Mismatches: 115
Query Match: 7.75% Indels: 91
DB: 9 Gaps: 10

US-09-889-756a-2 (1-412) x US-09-815-242-7049 (1-1173)

QY 75 ValArgAlaGlnValGlyGlyIleIleGlnLysArgLeuPheGlnGlySerTyrVal 94
DB 190 GTATCATCACAAGTTGACGATATCGAAGATTATGACAGCAATATGATTAAGTC 249
QY 95 ArgAlaGlyGlnProLeuTyrGlnIleAspSerSerThrTyrGlnAlaAsnLeuGlnSer 114
DB 250 CATCAGGTGATATCTTACGTGAATGATGATCAATGCGAATCTTGAATTCGACAG 309
QY 115 AlaArgAlaGlnLeuAlaThrAla 122
DB 310 GCCAAAGCAATCTGCCAATGCTGTACGTACGAGCAACCTGGGTTTCATCTGTCAG 369
QY 123 -----GlnAlaThrLeuAlaLysAlaAspAlaAspLeu 133
DB 370 CAATTGCAATCAGCTGTCAGCGCAATGAATTTCTTACGTCAAGCTCAAGGCAATTG 429
QY 134 AlaArgTyrLysProLeuValAlaAlaGlnAlaValSerArgGlnLysTyrAspAlaAla 153
DB 430 GCGGCGGAGTTCAACTGTAAGAAATGGGTGCGATTGTAAGATCTTTCACACGCA 489
QY 154 ValThrAlaLysArgSerAlaGlnAlaGlyValLysAlaAlaGln 168
DB 490 AAGAGAGCTGTCGACTTGCACAAAGCAATTTAAATGCTCTAAAAATCAATTGCTGCT 549
QY 168 ----- 168
DB 550 AATCAGCTTATTCGTAACGTTCCATTACGTGACCAACCGCAATATACAGATTCGATA 609
QY 169 AlaAlaIleLysSerAlaGlyIleAsnLeuAsnArgSerArgIleThrAlaProIleSer 188
DB 610 AACTCGCTTAAACAGGCTTGTTAAATTTGCAACGTACAAAAATTTAGACGCAATTGAT 669
QY 189 GlnPheIleGlnSerLysValSerGlnGlyThrLeuLeuAsnAlaGlyAspThrThr 208
DB 670 GGCATATGGCGCTCTTAATGTTCAAGTTGGGCAACCGTTTCAGTGGGTGGCGCTTA 729
QY 209 ValLeuAlaThrIleArgGlnThrAsnProMetTyrValAsnValThrGlnSerAlaSer 228
DB 730 ATGCGCATGCTATTAATGAACAAATGTGGCTAGACGTACTTCAAGAAACCAATTA 789
QY 229 GluValMetLysLeuArgArgGlnIle-----AlaGlnGlyLysLeu 243
DB 790 ACGAATATGCGCATTTGTCAGCCAGTAAATAATCATTTTATTAATCGGTAATAATAA 849
QY 244 AlaAlaAspGlyValIleAlaValGlyIleLysPheAspGlyThrValTyrProGln 263
DB 850 GAATTTATGCGCGATTAAT--GGTATTGAATGGGAGCATGCAATGCAATTT----- 900
QY 264 LysGlyArgLeuPheAlaAspProValValAsnGlnSerThrGlnIleThrLeu 283
DB 901 -----TCTCTTTTG-----CCTTCACAAATCTCTAGGGTAATCGATTAAGATG 945
QY 284 ArgAlaIleValProAsnAspGlnAsnIleLeuMetProGlyLeuTyrValArgValLeu 303

```

```

DB 946 GTGCAGCGGTACCT-----GTGCAATTAA 972
QY 304 MetAspGlnValAlaValAspAsnAlaPheValValProGlnGln----- 318
DB 973 TTACAT-----CCACAACATTTACAGAAACGCCA 1002
QY 319 --AlaValThrArgGlyAlaLysAspThrValMetIleValAsnAlaGlnGlyMet 337
DB 1003 TTGAGAATTGCTCTTTCGCAACGCAAAAGTTAAGATTTCAGATTATTCAGCGCAATG 1062
QY 338 -----GluProArg 340
DB 1063 CTACGAGAAAAAACAGAACCGAAA 1086

RESULT 13.
US-09-769-744A-93
Sequence 93, Application US/09769744A
Publication No. US20030134407A1
GENERAL INFORMATION:
APPLICANT: Le Page, Richard WF
APPLICANT: Wells, Jeremy M
APPLICANT: Hamifly, Sean B
APPLICANT: Hansbro, Philip M
TITLE OF INVENTION: Proteins
FILE REFERENCE: PWC/P211220
CURRENT FILING DATE: US/09/769, 744A
PCT/GB99/02452
PRIOR FILING DATE: 1999-07-27
PRIOR APPLICATION NUMBER: GB 9816336.3
PRIOR FILING DATE: 1998-07-27
PRIOR APPLICATION NUMBER: US 60/125329
NUMBER OF SEQ ID NOS: 196
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 93
LENGTH: 1200
TYPE: DNA
ORGANISM: Streptococcus pneumoniae
US-09-769-744A-93

Alignment Scores:
Pred. No.: 8.3e-07 Length: 1200
Score: 151.50 Matches: 98
Percent Similarity: 41.58% Conservative: 65
Best Local Similarity: 25.00% Mismatches: 165
Query Match: 7.50% Indels: 65
DB: 12 Gaps: 16

US-09-889-756a-2 (1-412) x US-09-769-744A-93 (1-1200)

QY 12 AlaAlaLeuAlaAlaAlaValAlaLeuValLeuSerSerCysGlyLysGlyLysAla 31
DB 43 GCAGCAATCGTCTGCGAGTAGTGTATG-----GGTCTGGGGGCAATTTTA 93
QY 32 -----AlaGlnGlyGlnProAlaGlyArgGlnAlaProAlaProValAlaGlyVal 49
DB 94 CTTTGAAGAACCTTCTCTCAGCTCTTAAGATGAGT-----ACTCATCTT 144
QY 50 ValThrValHisProGlnThrValAlaLeuThrValGlnLeuProGlyArgLeuGlnSer 69
DB 145 GTTGTGCCAAGAGAGAGAGCGCTCTCTCTGTTTATTTGTCAGGACAGTAACGCA 204
QY 70 LeuArgThrAlaAspValArgAlaGlnValGly--GlyIleIleGlnLysArgLeuPhe 88
DB 205 AAAAATGAACAAATATGTTTATTTTGAATGATGAGGTGATTTAGTAATCTTGT 264
QY 89 GlnGlnGlySerTyrValArgAlaGlyGlnProLeuTyrGlnIleAspSerSerThrTyr 108
DB 265 TCTGTGGCGATTAAGTACGAGGAGCGGCTTTTGTGTAAGTAAGTACGATTCAGAAAGCG 324
QY 109 GluAlaAsnLeuGlnSerAlaArgAlaGlnLeuAlaThrAla----- 122

```



```

Db 7419367 GCGCTTGAGATCGGGGGTGTGTAGCGGACGTCACCTACTCCATCTACGCTTGACCCCGGG 7419308
Qy 202 euAenAlaGlyAspThrValIleuAlaThrIleArgIleThrAspProwet----- 219
Db 7419307 CGAACCGTCACCCGGGGTTCACCGGGCCACGGTGAGAGGGCTACAGCCGGTTGCGCGTG 7419248
Qy 220 -----TyrV 221
Db 7419247 TGACCGCGCGACGAGACCGCGCGCTTGTGGCGGACCTGCTGCTCTCGGCGACCGGG 7419188
Qy 221 aAlaenValThrIleSerAlaSerGluValMetIleuArgIleGlnIleAlaGluIle 241
Db 7419187 TCCGAGTCCGCGACGATGCGCGCGCGCGCTCCGACGTAAGCGCTCGCGCGACGAC 7419128
Qy 241 yLeuLeuAlaAlaAspGlyValIleAlaValGlyIleLeuPhe-----AspAspG 258
Db 7419127 GCCGTCGCGATGCGGATGCGGCTGTCGAGTCCGCCCGGAAAGTGAAGTACCGACGACG 7419068
Qy 258 lYThrValTyrProGluIleGlyIleuLeuPheAlaAspProVal----- 273
Db 7419067 CCGCGTACAGCCCGCGCGGAGCGGCTCCAGTTCGTCGATGATCGAGGCGCGGGGC 7419008
Qy 274 -----ValAsnGluSerThrGlyIleThrIleuArgAlaAlaVal----- 287
Db 7419007 TTGGGCGCGCGGAGAGGGTCCGCGCGGAGACGCGCTCAGACGTCGAGCGCGTA 7418948
Qy 288 -----ProAsnAspGlnAsnIleLeuMetProGly-----LeuTyrV 300
Db 7418947 CGCCCCCGCGGACCGC-----CCCGTTCACGCTCGACAGATGTCATCAGC 7418900
Qy 300 aAlaArgValLeuMetAspGlnValAlaValAspAsnAlaPheValValProGlnIleAlaVal 320
Db 7418899 TGGCAGTAGCGCTCGATGTCATGAGTGCAGCAGC----- 7418865
Qy 320 aIThrArgIleAlaLeuAspThrValMetIleValAsnAlaGlnIleGlyIleMetGluProA 340
Db 7418864 -----CTCGACGGAACCGGGCTCGCACCCCGC 7418837
Qy 340 rGluValThrValAlaGlnGlnIleGlnIleThrAsnTrpIleValThrSerGlyIleuIleu 360
Db 7418836 CCCGAGATC---GTTCGCGCCCGACGATCGACGATCAGTGTCTGGCGCTCTTGGCG 7418780
Qy 360 spGluAspIleValValAlaGluIleIle-SerIleAlaGlyIleThrGlyAlaIleLeu 379
Db 7418779 TCGGCGCAGACGCTCTCGCGAGGGCTGTCTCTCGCGGGGTGCGCGCGCGCGCGC 7418720
Qy 380 ValThrProIleGluTrpAlaSerSerGluAsnGlnAlaAlaAlaProGlnSerGlyVal 399
Db 7418719 GTGCCCGCGATGGGGTGGACC-----ATCGCTGCGCGTCTCGACCTTG 7418675
Qy 400 GluThrAlaSer 403
Db 7418674 ACAGGGGCTCG 7418663

RESULT 15
US-09-741-669-242
; Sequence 242, Application US/09741669
; Patent No. US20020022718A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlson, Karl L.
; APPLICANT: Zykkind, Judith W.
; TITLE OF INVENTION: Genes identified as required for
; TITLE OF INVENTION: proliferation of E. coli
; FILE REFERENCE: ELITRA.009A
; CURRENT APPLICATION NUMBER: US/09/741.669
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 60/173005
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 481
; SOFTWARE: PaasSEQ for Windows Version 4.0
; SEQ ID NO 242
; LENGTH: 1068

```

```

; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1068)
US-09-741-669-242

Alignment Scores:
Pred. No.: 1 79e-05 Length: 1068
Score: 138.50 Matches: 52
Percent Similarity: 42.42% Conservative: 32
Best Local Similarity: 26.26% Mismatches: 55
Query Match: 6.86% Indels: 59
DB: Gaps: 6

US-09-889-756a-2 (1-412) x US-09-741-669-242 (1-1068)

Qy 65 GlyArgLeuGluSerLeuArgThrAlaAspValArgAlaGlnValGlyIleIleGln 84
Db 121 GGGCGCATTTGAAACGACGAA---GTGGATATTGCCAGCAAAATTGCCGGCGTATCGAC 177
Qy 85 LysArgLeuPheGlnGluIleSerIleValArgAlaGlyIleProIleuTyrGlnIleAsp 104
Db 178 ACAATTCTGTGAAGAAGCAAGTTGTCGGAAGTGAAGTGTGCGGAGATGAT 237
Qy 105 SerSerThrTyrGluAlaAsnLeuGluSerAlaArgAlaGlnLeuAlaThrAlaGlnAla 124
Db 238 ACTCGCGTGTTCGAGAAACGACGATCGGAAGCCATCGCGCAATCAAGAAGCACAAGC 297
Qy 125 ThrLeuAlaLysAlaAspAlaAspLeu----- 133
Db 298 GCCGTGTCGCCGCGAGGCTTGTTCGAGCAACGAAAGCAAACTGTCGCCGACAG 357
Qy 134 -----AlaArgTyr 136
Db 358 TCGCTGTATATCAACGCCGAGCACTGACTCCGTACCAAAAGCTCATACGCTTCC 417
Qy 137 LysProLeuValAlaAlaGluAlaValSerArgGlnIleuTyrAsp----- 151
Db 418 GCTTCACTGCCCCAAGAGGGGCTATTTCGCCCAACAGCTGTGATGATCGCGCGCC 477
Qy 152 -----AlaAlaValThrAlaLysSerAlaGluAlaGlyValIleLysAlaAlaGlnAla 169
Db 478 GCTGAGCGCGCGAGCTGCTGGAATGCGGAAAGCTAGATTCGCTTCTAAAGC 537
Qy 170 AlaIleLysSerAlaGlyIleAsn----- 177
Db 538 GCTATGAAAGCGGACGACCAATATCATTCAGGCGCAACCCGCTGGAAGCGCACAA 597
Qy 178 -----LeuAsnArgSerArgIleThrAlaProIleSer 188
Db 598 GCCACTGAACGGCGCATTCGCCGCAATATGATGACGACGCACTGAAGCCCGCGTAC 657
Qy 189 GlyPheIleGlyIleSerIleValSerGlu---GlyThrLeuLeuAsnAlaGly 205
Db 658 GGAAGCGGTG---CAGTATCGGGTGTGCGAGCCAGGGAAGTGTGCGCGCAGGC 708

RESULT 16
US-09-855-754-1/c
; Sequence 1, Application US/0985754
; Publication No. US20020192237A1
; GENERAL INFORMATION:
; APPLICANT: BOURSAUX-BUPE, CAROLINE
; APPLICANT: GUISO-MACLOUF, NICOLE
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,
; TITLE OF INVENTION: BORDETELLA PARAPERTUSSIS, AND BORDETELLA
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
; FILE REFERENCE: 03495-0206-00000
; CURRENT APPLICATION NUMBER: US/09/855, 754
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/206,969

```

; PRIOR FILING DATE: 2000-05-25
 ; NUMBER OF SEQ ID NOS: 24
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 1
 ; LENGTH: 3000
 ; TYPE: DNA
 ; ORGANISM: Bordetella bronchiseptica
 ; US-09-885-754-1

Alignment Scores:

Pred. No.:	0.0002	length:	3000
Score:	134.50	Matches:	115
Percent Similarity:	37.29%	Conservative:	42
Best Local Similarity:	27.32%	Mismatches:	171
Query Match:	6.66%	Indels:	94
DB:	10	Gaps:	17

US-09-889-756a-2 (1-412) x US-09-885-754-1 (1-3000)

```

QY 9 MetArgAlaAlaAlaLeuAlaAlaAlaValAlaLeuSerSerCysGlyTyrGly 28
DB 2298 ATGCAGCTGTCGTGGTGGCGCCGCTGCGCC-----GCTAAAGCC 2257
QY 29 GlyAspAlaAlaGln---GlyGlyGlnProAlaGlyArgGluAlaProAla--- 46
DB 2256 GCGTGGCGCGCGGTATGAGCCGCGCAGCCGCC-----CAGTCCAGCGCCGCCCGCC 2203
QY 47 -----ValGlyValAlaThrValHisProGlnThrAlaAlaLeuThrValGluLeuPro 64
DB 2202 CACGCCACCGCGTGGTGGCGCCCGCAGCTCGAAGCGCGCCACCTTCTGTGCAAGCGCG 2143
QY 65 Gly-----ArgLeuGluSerLeuArgThrAlaAspValArgAlaGlnValGly 80
DB 2142 CCCGCGCGGTTGTCAGTGTGCTGGCGTGGCGAAGCGCGCGCCCAAGCGCGCGCCG 2083
QY 81 GlyIleIleGlnLysArgLeuPheGlnGlnGlySerTyrValArgAlaGlyGlnProLeu 100
DB 2082 GTCCGAGATTCAAGCCCACTC-----GCCCAGCGCGCTT 2050
QY 101 TyrGlnIleAspSer-SerThrTyrGluAlaAsnLeuGluSerAlaArgAlaGlnLeuAl 120
DB 2049 GACACACCATTCCTTTCGCGCTACCAAGAGCTGCTG-----GCCAGGCCACCCACC 1996
QY 120 aThrAlaGlnAlaThrLeuAlaAlaAlaAspAlaAspLeuAlaArgTyrLys-ProLeu 140
DB 1995 CGTGTGACCGCGCGTGGCGCGCGGCGAACAATCCCTG-----CCCGCGCGCGG 1945
QY 140 AlaAlaAlaGlnAlaValSerArgGlnGluTyrAspAlaAlaValThrAlaLysSerAla 160
DB 1944 TTGCGCGCGCGCGCTTCGCGCTGCTGTG--GCGGCTGCGCGCGCTGCGCGCGCTG 1888
QY 160 IagLualaGlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyLysLeuAsnLeuAsn 180
DB 1887 CCGCGCGCTGGGACCGCGCTGGGACCGCGCGCGCGCGCGCTGGGCGCGCGCGCGCG 1828
QY 180 rGserArg-IleThrAlaProIleSerGlyPheIleGlyGlnSerLysValSerGlnGly 199
DB 1827 CTTGCGCGCGACCAAGCTCCACTGCCCATTTGGCGCGCGCATTCATGACGGTAGCT 1768
QY 200 ThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAsnProMet 219
DB 1767 ACCGATATC-----GACCTTGCCGCTCTTGTGGCAAGGTAAGGTGCG----- 1723
QY 220 TyrValaAsnValThrGlnSerAlaSerGluValMetLysLeuArgArgGlnIleAlaGlu 239
DB 1722 -----CGCGTGCCTCGTGGCGCTGCAACGACGACGTGTTGCGCT 1678
QY 240 GlyLysLeuLeuAlaAla-----AspGlyValIleAlaValGlyIleLysPhe 255
DB 1677 GCGCGGCTGCGTGGCTGTGGCGGACCAACAGCCTGTGCTGGCGCGCTGCGCGCAT 1618
QY 256 AspAspGlyThrValTyrProGluLysGlyArgLeuPheAlaAspProValValAsn 275
  
```

```

DB 1617 GACGAC-----CAGCTTGTGCTGACCCCGCAGCTCCG 1585
QY 276 GluSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeuMet 295
DB 1584 GAAACATTCATTCGGGAACAGACCCCGAACCAGCGGTATCGACCATCAGGCACTTGA 1525
QY 296 ProGlyLeuTyrValArgValLeuMetAspGlnValAlaValaAspAsnAlaPheValVal 315
DB 1524 CCGCCAGCTTCGCGCGCTGCTGGAATCGACCGCTCGCTGCGCCAG----- 1474
QY 316 ProGlnGlnAlaValThrArgGlyAlaLysAspThrValMetIleValAsnAlaGlnGly 335
DB 1473 CCGAGCGCGCGC-----GACGTTCAAGTGTTCGTCATGACCCA 1435
QY 336 GlyMetGluProArgGluValThrValAlaGlnGlnGlnGlnThrAsnTrpIleValThr 355
DB 1434 GGT----- 1432
QY 356 SerGlyLeuLysAspGlyAspLysValValAlaGlnGlyLysSerIleAlaGlyIleThr 375
DB 1431 ---GCGTGTGATGACAGCGAGTCGACCGCGCGGTAGCGCCGTCATCGGG-CCT 1376
QY 376 Gly-----AlaLysValThrProLysGluTyrPalasSerSerGlu 389
DB 1375 GCGTGGCCAGCGCCACCTCGACCGCGCGCTCGACGCGCTGATGGAGGACAGCTCCG 1316
QY 390 AsnGlnAlaAlaProGlnSerGlyValGlnThrAlaSerGlnAlaLysThrAla 408
DB 1315 TCGGAGCATGTCCCTGCGCCCTGCGCGCGCGCGCGCGCGCGCGCGCGCTTCAAGGCT 1259
  
```

RESULT 17

US-10-227-353-3/c
 ; Sequence 3, Application US/10227353
 ; Publication No. US20030087383A1

GENERAL INFORMATION:

APPLICANT: CLARE, JEFFREY J.
 ROMANOS, MICHAEL A.

TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN YEAST

NUMBER OF SEQUENCES: 17

CORRESPONDENCE ADDRESS:
 ADDRESSEE: Millen, White, Zelano & Branigan, P. C.
 STREET: 2200 Clarendon Blvd., Suite 1400
 CITY: ARLINGTON
 STATE: VA
 COUNTRY: USA

ZIP: 22201

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/227,353
 FILING DATE: 26-Aug-2002

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/460,269C
 FILING DATE: 02-Jun-1995

ATTORNEY/AGENT INFORMATION:
 NAME: Lebovitz, Richard M.
 REGISTRATION NUMBER: 37,067
 REFERENCE/DOCKET NUMBER: Popov-2

TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 243-6333
 TELEFAX: (703) 243-6410

INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3000 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA
 FEATURE:


```

Db 2687 CCTCGCCCGGAGATCCCGCTGACGCCGAGACACCGCCCGGAGTCCGAGCGCG 2628
Qy ArGThirAlaaspValArgAlaGlnValGlyIleIleGlnIlyArgLeuPheGlnIlu 90
Db 2627 AAGACGAGTTGCGCCGTCGCCCGCTCCAGCGGCTGTCACCGCG-----TCGCC 2580
Qy 91 GlySerTyValArgAlaGlnIleProLeuTyGlnIleAspSerSerThyTyrGlnAla 110
Db 2579 GACACCTTCGTACGCTGACGACGATCTCGGTGTTCCGCTGCTCGTCCGAACTGC----- 2529
Qy 111 AsnLeuGlnSerAlaArgAlaGlnLeuAlaThraGlnAlaThraLeuAlaGlnAlaasp 130
Db 2528 -----GCCGCTTCCTAGCGCCAGCGCGACGTCAGCTACGCGGGCGCGCGCTCGCGGTC 2475
Qy 131 AlaAspLeuAlaArgTyTyrLeuProLeuValAlaAlaGlnAlaValaSerArgGlnIluTyr 150
Db 2474 AGCTGATCTGTCGACGACGCGGATGTCCTCGCTCGACGCGCGACCGACGAGATTCTGAC 2415
Qy 151 AspAlaAlaValThraAlaLysArg-----SerAlaGlu 161
Db 2414 GACCCCGGCTCCACCCGTCGCGCGCGCGCGACGTCGCCAGAACGCGAGCGCGCGCGAC 2355
Qy 162 AlaGlyVallyAlaAlaGlnAlaAlaIleLysSerAla-----Gly 175
Db 2354 GCGAGTGGGAGGCGGACGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2295
Qy 176 IleAsnLeuAsnArgSerArg-----IleThraIleProIleSerGlyPheIle 191
Db 2294 GCCAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2235
Qy 192 GlyIleSerLysValSerGluGlyIleThraLeuAsnAlaGlyAspThr---ThraLeu 210
Db 2234 ACCGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 2175
Qy 211 AlaThrIleArgGlnThraAsnProMetCtyrValaAsnValThraGlnSerAlaSerGluVal 230
Db 2174 TCGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2137
Qy 231 MetLysLeuArgArgGlnIleAlaGlnIleLysLeuLeuAlaAlaAspGlyValIleAla 250
Db 2136 ---CGCGTACACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2110
Qy 251 ValGlyIleLysPheAspAspGlyThraValTyrProGluLysGlyArgLeuLeuPheAla 270
Db 2109 -----GAAACAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2065
Qy 271 AspProValValaAsnGlnSerThryGlyIleIleThraLeuArgAlaAlaValaProAsnAsp 290
Db 2064 GATCAGCTGCTAGTCTGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2026
Qy 291 GlnAsnIleLeuMetProGlyLeuTyrValaArgValaLeuMetAspGlnValaAlaValaAsp 310
Db 2025 -----GTCCGCGTACCAAGTCTG-----CGGAGCGCGCGCGCGCGCG 1993
Qy 311 AsnAlaPheValaValaProGlnAlaAlaValaThraArgIlyAlaLysAspThrValaMetIle 330
Db 1992 GGGGAGAGCGTCCCGGACAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1933
Qy 331 ValaAsnAlaGlnIlyGlyMetGlnProArgGlnValaThraValaAla----- 345
Db 1932 CGCGGTCACAGAGCGCGCG-----CGGAGGTCTGCGGTGTCACCGCGTACGCGGTA 1882
Qy 346 -----GlnGlnGlnGly 349
Db 1881 CGCCGAGACAGAGACAGAGCG 1861

```

RESULT 21

```

US-10-227-353-5/c
; Sequence 5, Application US/10227353
; Publication No. US20030087383A1
; GENERAL INFORMATION:
; APPLICANT: CLARE, JEFFREY J.
; ROMANOS, MICHAEL A.

```

```

/ TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN
/ YEAST
/ NUMBER OF SEQUENCES: 17
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Millen, White, Zelano & Branigan, P.C.
/ STREET: 2200 Clarendon Blvd., Suite 1400
/ CITY: ARLINGTON
/ STATE: VA
/ COUNTRY: USA
/ ZIP: 22201
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/10/227,353
/ FILING DATE: 26-Aug-2002
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US/08/460,269C
/ FILING DATE: 02-Jul-1995
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Lebovitz, Richard M.
/ REGISTRATION NUMBER: 37,067
/ REFERENCE/DOCKET NUMBER: Popov-2
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (703) 243-6333
/ TELEFAX: (703) 243-6410
/ INFORMATION FOR SEQ ID NO: 5:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 3000 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 145..2910
/ SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-10-227-353-5
Alignment Scores:
Pred. No.: 0.000421 Length: 3000
Score: 131.50 Matches: 114
Percent Similarity: 37.35% Conservative: 44
Best Local Similarity: 26.95% Mismatches: 174
Query Match: 6.51% Indels: 92
DB: 14 Gaps: 16
US-09-889-756a-2 (1-412) x US-10-227-353-5 (1-3000)
Qy 9 MetArgAlaAlaAlaLeuAlaAlaValaLeuValaLeuSerSerCysGlyIlyGly 28
Db 2231 ATGACCGCTGTGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2290
Qy 29 GlyAspAlaAlaGln---GlyGlyGlnProAlaGlyArgGlnAlaProAlaProVal--- 46
Db 2289 GCGGTCCGCGCGCGGTATAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2236
Qy 47 -----ValGlyValaIleThraIleProGlnThraValaAlaLeuThraValaGlnLeuPro 64
Db 2235 CACCGCACCGCGGTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2176
Qy 65 Gly-----ArgLeuGlnSerLeuArgThraAlaAspValaArgAlaGlnAlaGly 80
Db 2175 CCCGCGCGGTTCACGATTCCTGCTGCGTGGCGAACCGCGCGCGCGCGCGCGCG 2116
Qy 81 GlyIleIleGlnIlyAsnArgLeuPheGlnGlnGlySerTyrValaArgAlaGlyGlnProLeu 100
Db 2115 GTCCGATTCAGGGCAACTC-----GCCACAGCGCTT 2083
Qy 101 TyrGlnIleAspSer---SerThyTyrGlnAlaAsnLeuGlnSerAlaArgAlaGlnLeuAl 120

```

[illegible]

RESULT 22
US-09-855-754-3/c
Sequence 3, Application US/09855754
Publication No. US20020192237A1
GENERAL INFORMATION:
APPLICANT: BOURSAUX-EUDE, CAROLINE
APPLICANT: GUISSO-MACLOUF, NICOLE

```

? TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
? TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORETTELLA PERTUSIS,
? TITLE OF INVENTION: BORETTELLA PARAPERTUSIS, AND BORETTELLA
? TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
? TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
? FILE REFERENCE: 03495-0206-00000
? CURRENT APPLICATION NUMBER: US/09/855,754
? CURRENT FILING DATE: 2001-09-10
? PRIOR APPLICATION NUMBER: 60/206,969
? PRIOR FILING DATE: 2000-05-25
? NUMBER OF SEQ ID NOS: 24
? SOFTWARE: PatentIn Ver. 2.1
? SEQ ID NO 3
? LENGTH: 3116
? TYPE: DNA
? ORGANISM: Bordetella parapertussis
US-09-855-754-3

Alignment Scores:
Pred. No.: 0.00044 Length: 3116
Score: 131.50 Matches: 114
Percent Similarity: 37.35% Conserved: 44
Best Local Similarity: 26.95% Mismatches: 174
Query Match: 6.51% Indels: 92
DB: 10 Gaps: 16

US-09-889-756A-2 (1-412) x US-09-855-754-3 (1-3116)
QY 9 MetArGaAlaAlaLeuAlaAlaAlaValAlaLeuValLeuSerSerCyGlyGly 28
Db 2331 ATGCACCGCTTCGGGTGGCCGCCGCCGTCGCC-----GTTAAAGCC 2290
QY 29 GlyAspAlaAlaGln---GlyGlyGlnProAlaGlyArgGluAlaProAlaProVal--- 46
Db 2289 GCGGTCCGCCGCCGTATAGCCGCCCAAGCCGCC-----CAGGTGCAGCCGCCGCCGCC 2236
QY 47 -----ValGlyValValThrValHisProGlnThrValAlaLeuThrValGluLeuPro 64
Db 2235 CACC GCCCACCGCGTGTCGCCGCCCAAGCTCGAACCCGCCCACTTCTGTCGAAGCCGC 2176
QY 65 Gly-----ArgLeuGlnSerLeuArgThrAlaAspValArgAlaGlnValGly 80
Db 2175 CCCGCCCGGTTGTCACAGTTGTCGCGGTGGCCAAACCCGCCGCCCAAGGCCGCCGCC 2116
QY 81 GlyIleIleGlnLysArgLeuPheGlnGluGlySerTyValArgAlaGlyGlnProLeu 100
Db 2115 GTCCGCAATTCAGGGCCAATC-----GCCAAGCGCTT 2083
QY 101 TyrGlnIleAspSer-SerThrTyrgluAlaAsnLeuGlnSerAlaArgAlaGlnLeuAl 120
Db 2082 GGACAACGCATTCCTTCGCGGTCACAGACGTCGTG-----GCCAAGGCCCAATCCACC 2029
QY 120 aThraAlaGlnAlaThrLeuAlaLysAlaAspAlaAspLeu-----AlaArgTyrIrp 138
Db 2028 CGTGTTCACCCGCCCGTGGCGCGGGGAGCAACTCCCTCCCGCGCGCGTTGCCGCC 1969
QY 138 roLeuValAlaAlaGluAlaValSerArgGlnGluTyArgAspAlaAlaValThrAlaLys 158
Db 1968 CGGCGCTTCGCCGCTCTCTGTGGCGGCTCGCGCGGCTCGCGCGGCTCGCGGCTCGCG 1909
QY 158 rGserAlaGluAlaGlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsn 178
Db 1908 CGGCTGCGGCGGCTGGGAGCCGGGCTGGGAGCCGGCTGGGCGCGGCTTGGGCGCGG 1849
QY 178 euSnArGserArg-IleThrAlaProIleSerGlyPheIleGlyGlnSerLysValSer 197
Db 1848 CGGCGCTTCGCCGCCCAAGGCTCCCATTCGCAATTCGCGCGGCGGCAATCGATGCG 1789
QY 198 GluGlyThrLeuLeuAlaGlyLysArgThrThrValLeuAlaThrIleArgGlnThrAsn 217
Db 1788 GTAGGTACCAATTC-----GACCTTCGCCGCTCTTGTGGCAAAGGTAAAGTCCG 1738
QY 218 ProMetTyValAsnValThrglnSerAlaSerGluValMetLysLeuArgGlnIle 237

```

```

Db 1737 ----- :CGCGCTGCTGTGTGGCGTCTGCACACAGATAGTGT 1698
OY 238 AlaGlugLyLysLeuLeuAlaAa-----AspGlyValIleAlaValAllyle 253
Db 1698 GCGCGTGGCGCGGCTGTGCTGTGTGGCGGAGCACACAGCGCTGTGCGCGTGTGGCGTGC 1638
OY 254 LysPheAspAspGlyThrValTyPrProGluLysGlyArgLeuLeuPheAlaAspProVal 273
Db 1638 CCGCATGACGAC-----CAGCTTGTGCTTACGCGCCAG 1606
OY 274 ValAsnGluSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAspGlnAsnIle 293
Db 1605 GTCCGCGAAGACATTCATCGCGAAGACGCCGACCGCCGACCGTATGACATCGATCAGGAC 1546
OY 294 LeuMetProGlyLeuTyRValArgValLeuMetAspGlnValAlaValAspAsnAlaPhe 313
Db 1545 CTTGAAACCGCCAGCTTGTGCGCGCGTGTGGAATCGACGCTGCGCTGCGCCAG--- 1488
OY 314 ValValProGlnGlnAlaValThrArgGlyAlaLysAspThrValMetIleValAsnAla 333
Db 1488 -----CGCAGCGCGCC-----GACGTTCGAGTTGTCCGTCA 1456
OY 334 GlnGlyGlyMetGluProArgGluValThrValAlaGlnGlnGlnGlyThrAsnTrpIle 353
Db 1455 GACCCAGGT----- 1444
OY 354 ValThrSerGlyLeuLysAspGlyAspLysValAlaValGluGlyIleSerIleAlaGly 373
Db 1446 -----GGGTGTGTGATGAGACAGGAGTCGACCGCGGGTAGCGCCGTCCATCG 1396
OY 374 IleThrGly-----AlaLysLysValThrProLysGluThrPalaSer 387
Db 1395 GG-CCTGGCTGGCGCAGCGCCAGCTGACGCGCGCCCTGCGAGCGCTGGAAGGAGCA 1337
OY 388 SerGluAsnGlnAlaAlaAlaProGlnSerGlyValGlnThrAlaSerGluAlaLysThr 407
Db 1336 GCTCGGTGCGACGATGTGCGCTTCCCTTGGCGCGCGCCGCGCAGGTCACTTCAACG 1277
OY 408 Ala 408
Db 1276 GCT 1274

RESULT 23
US-10-114-170-260/c
: Sequence 260, Application US/10114170
: Publication No. US20030023075A1
GENERAL INFORMATION:
APPLICANT: Blattner, Frederick R.
Burland, Valerie
Perna, Nicole T.
Plunkett, Guy
Melich, Rod
TITLE OF INVENTION: No. US20030023075A1 Sequences of E. coli O157
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESS: Charles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/114,170
FILING DATE: 01-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION NUMBER: 09/453,702

```

```

? FILING DATE: 03-DEC-1999
? APPLICATION NUMBER: 60/110,955
? FILING DATE: 04-DEC-1998
? ATTORNEY/AGENT INFORMATION:
? NAME: Seay, Nicholas J.
? REGISTRATION NUMBER: 27386
? REFERENCE//DOCKET NUMBER: 960296.95017
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (608) 251-5000
? TELEFAX: (608) 251-9166
? INFORMATION FOR SEQ ID NO: 260:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2398 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? SEQUENCE DESCRIPTION: SEQ ID NO: 260:
US-10-114-170-260

Alignment Scores:
Pred. No.: 0.000397 Length: 2398
Score: 130.50 Matches: 61
Percent Similarity: 40.28% Conservative: 55
Best Local Similarity: 21.18% Mismatches: 85
Query Match: 6.46% Indels: 87
DB: 14 Gaps: 9

US-09-889-756A-2 (1-412) x US-10-114-170-260 (1-2398)
QY 60 ThrValGluLeuPProGlyArgLeuGluSerLeuArgThrAlaAspValArgAlaGlnVal 79
| | | | | | | | | | | | : : : : : : : : : : : :
DB 2349 ACGGTAGAACCTTCCGGCGCGGAACACAGTGTCAGTCGCTGAA----- 2305
      80 GlyIleIleIleGlnLysArgLeuPheGlnGlnGlySerTyValArgAlaGlyGlnPro 99
      |||||:::||||:::||||:::||||:::||||:::||||:::||||:::||||
DB 2304 GGCGGAATTCTGATCATCTCGACGTGAAGTTGGCATTATTCGAACAGGCCGACGCCG 2245
QY 100 LeuTyrgInlleaspSerSerThrTyrgInlAlaasnLeuGluSerAlaArgAlaGlnLeu 119
      |||||:::||||:::||||:::||||:::||||:::||||:::||||
DB 2244 CTGGCGCAGGCTTAACCCACCACCAACCGGAATCCACGTCGAGMAAGCATGTCGCCTG 2185
QY 120 AlaThrAlaGlnAlaThrLeuAlaValAlaAspAlaAspLeuAlaArgTyrlAspProLeu 139
      |||||:::||||:::||||:::||||:::||||:::||||:::||||
DB 2184 TACCGCGCGCTGGGAACCTCGCGCGCGCTGCGGCGGAGAgtGACAAT--AAACGCTG 2128
      |||||:::||||:::||||:::||||:::||||:::||||:::||||
QY 140 ValAlaAlaGlnAlaValAserArgGlnGluTyrlAspAlaAlaValAlaThrAlaArgSer 159
      |||||:::||||:::||||:::||||:::||||:::||||:::||||
DB 2127 GTCTTCCTCGTAGCAACGTGAATAA-----TCCCAGAACTTATTCGAGTGC- 2083
      |||||:::||||:::||||:::||||:::||||:::||||:::||||
QY 160 AlacIuaIagIyValLySAlaAlaGlnAlaAlaIlelysSerAlaGlyIleAsnLeuasn 179
      |||||:::||||:::||||:::||||:::||||:::||||:::||||
DB 2082 -----GAAACGGGGTGTACMACACTCGC 2059
QY 180 ArgSerArglleThrAlaProIleSerGlyPheIleGlyGlnSerLysValserGluGly 199
      |||||:::||||:::||||:::||||:::||||:::||||:::||||
DB 2058 CGGCAC-----GGGTGAACAAAGCCACACCGGG 2029
QY 200 ThrleuLeuasnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAsnPromet 219
      |||||:::||||:::||||:::||||:::||||:::||||:::||||
DB 2028 -----CTGACGCGAGGGGATTAGCTGTGTCATCGCAACtCGGAGATGACCCACCGCTG 1975
QY 220 TyrlValasnValThrGlnSerAlaSerGlyValMetLysLeuArgArgGlnIleAlaGlu 239
      |||||:::||||:::||||:::||||:::||||:::||||:::||||
DB 1974 GTAATAACAGGGCGCAGCAGCAGCGTCGAAGTACTCGCGCTTCGCAACCCAGGCCAATGAA 1915
QY 240 GlyLysLeuLeuAlaAlaAspGlyValIleAlaValGlyIleLysPheAspAspGlyThr 259
      |||||:::||||:::||||:::||||:::||||:::||||:::||||
DB 1915 ----- 1915
QY 260 ValTyrrProGluLysGlyArgLeuLeuPheAlaAspProValValasnGluSerThrcly 279
      |||||:::||||:::||||:::||||:::||||:::||||:::||||
DB 1915 ----- 1915

```


QY 280 GlnlleThrLeuArgAlaIaValProAsnAspGlnAsnIleLeuMetProGlyLeuTyr 299
 Db 1914 -----CTGGAACAACTCGACGCGACGCACTACTAC 1876
 QY 300 ValArgValLeuMetAspGlnValAlaValAspAsnAlaPheValValProGlnGlnAla 319
 Db 1875 GTTACAGCGC---CGGAAAGATTAGCAAAAGCCAAATGCGAAGTGGAAACCCACGTTCC 1819
 QY 320 ValThrArgGlyAlaIaValAspThrValMetIleValAsn-----AlaGlnGly 335
 Db 1818 GTGATCCGTGACCGGAAGACTCTCTACCAAGCTCAACTTACCGCGCGGTGAGGC 1759
 QY 336 GlyMetGluProArgGluValThr 343
 Db 1758 ATTGTCAGATATTGATGTCACC 1735
 RESULT 24
 US-10-114-170-39
 ; Sequence 39, Application US/10114170
 ; Publication No. US20030023075A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Blatner, Frederick R.
 ; Perna, Nicole T.
 ; Plunkett, Guy
 ; TITLE OF INVENTION: No. US20030023075A1 Sequences of E. coli O157
 ; NUMBER OF SEQUENCES: 265
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Quarles & Brady
 ; STREET: 1 South Pinckney Street
 ; CITY: Madison
 ; STATE: WI
 ; COUNTRY: US
 ; ZIP: 53701-2113
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Word Perfect 8.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/114,170
 ; FILING DATE: 01-Apr-2002
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 09/453,702
 ; FILING DATE: 03-DEC-1999
 ; APPLICATION NUMBER: 60/110,955
 ; FILING DATE: 04-DEC-1998
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Seay, Nicholas J.
 ; REGISTRATION NUMBER: 27386
 ; REFERENCE/DOCKET NUMBER: 960296,95017
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (608) 251-5000
 ; TELEFAX: (608) 251-9166
 ; INFORMATION FOR SEQ ID NO: 39:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 25165
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 39:
 US-10-114-170-39
 Alignment Scores:
 Pred. No.: 0.01 Length: 25165
 Score: 130.50 Matches: 61
 Percent Similarity: 40.28% Conservative: 55
 Best Local Similarity: 21.18% Mismatches: 85
 Query Match: 6.46% Indels: 87

DB: 14 Gaps: 9
 US-09-889-756a-2 (1-412) x US-10-114-170-39 (1-25165)
 QY 60 ThrValGluLeuProGlyArgLeuGlnSerLeuArgThrAlaAspValArgAlaGlnVal 79
 Db 24240 ACCGTGAACCTTCCGGGCGCGAAGGTGGTGCAGTGGCTGAA----- 24284
 QY 80 GlyGlyIleIleGlnIlySarGleuPheGlnGlySerTyrValArgAlaGlyGlnPro 99
 Db 24285 GCGGGAATTCGTATCATCTCCAGCGTAAGATTGGCGATTTGTCCAGACGGCCAGCCG 24344
 QY 100 LeuTyrGlnIleAspSerSerThrTyrGluAlaAsnLeuGlnSerAlaArgAlaGlnLeu 119
 Db 24345 CTGCGCAGCTTAACCGCAACCAAGCAATCCGATCCAGAGATGATGCGCGCTG 24404
 QY 120 AlaThrAlaGlnAlaThrLeuAlaIlySarAlaAspAlaAspLeuAlaArgTyrLysProLeu 139
 Db 24405 TACGCGCGCTGGCAACTCGCGCGCTCGCGCGGAGTGAAT--AAACGCTG 24461
 QY 140 ValAlaAlaGluAlaValSerArgGlnGluTyrAspAlaAlaValThrAlaIlySarSer 159
 Db 24462 GTCTTCCCTGACCACTGAATAA-----TTCCCGCACTTTCAGTGC----- 24506
 QY 160 AlaGluAlaGlyValIlySarAlaAlaGlnAlaIleLysSerAlaGlyIleAsnLeuAsn 179
 Db 24507 -----GAAACGCGTTGTACACACTCCG 24530
 QY 180 ArgSerArgIleThrAlaProIleSerGlyPheIleGlyGlnSerTyrValSerGlyGly 199
 Db 24531 CGCGAC-----GGTTGAACAAAGCCACACCGCGG 24560
 QY 200 ThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAsnPromet 219
 Db 24561 -----CTGACGAGGAGATTACCTCGTCAATCGCAACTGCGCATGACCCGCGCTG 24614
 QY 220 TyrValAsnValThrGlnSerAlaSerGluValMetLysLeuArgArgGlnIleAlaGlu 239
 Db 24615 GTAAACAGGGGCGACGACGACGCGTGAAGTACTGCGCTGACAGCCAGCGCATGAA 24674
 QY 240 GlyLysLeuLeuAlaAlaAspGlyValIleAlaValGlyIleLysPheAspAspGlyThr 259
 Db 24674 ----- 24674
 QY 260 ValTyrProGlnLysGlyArgLeuLeuPheAlaAspProValValAsnGlnSerThrGly 279
 Db 24674 ----- 24674
 QY 280 GlnlleThrLeuArgAlaIaValProAsnAspGlnAsnIleLeuMetProGlyLeuTyr 299
 Db 24675 -----CTGGAACAACTCGACGCGACGCACTACTAC 24713
 QY 300 ValArgValLeuMetAspGlnValAlaValAspAsnAlaPheValValProGlnGlnAla 319
 Db 24714 GTTACAGCGC---CGGAAAGATTAGCAAAAGCCAAATGCGAAGTGGAAACCCACGCTTCG 24770
 QY 320 ValThrArgGlyAlaIaValAspThrValMetIleValAsn-----AlaGlnGly 335
 Db 24771 GTGATCCGTGACCGGAAGACTCTCTACCAAGCTCAACTTACCGCGCGGTGAGGC 24830
 QY 336 GlyMetGluProArgGluValThr 343
 Db 24831 ATTGTCAGATATTGATGTCACC 24854
 RESULT 25
 US-09-938-842A-176
 ; Sequence 176, Application US/09938842A
 ; Patent No. US20020160378A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Harper, Jeff
 ; APPLICANT: Kreps, Joel
 ; APPLICANT: Wang, Xun
 ; APPLICANT: Zhu, Tong

Db 829 ATGAAGAAGAGTGTCTCCAAAGCAGAAAGAGAGAGCAAGTACTCTCGAAGCTTTGAAA 888
 Oy 223 ValThrGlnSerAlaSerGluValMetLysLeuArgGlnIleAlaGluGlyLysLeu 242
 Db 889 CGTCTCAAAAGAGTAAGAGATGATACAAAATAATGACAGAGAAAGTCGGAAATGCTGCTC 948
 Oy 243 LeuAlaAlaSerGlyValIleAlaValGlyLysPheSerPheSerPheGlyThrValTyrPro 262
 Db 949 -----TCGTGGTGTGTTTAAAGTTCTCGAATTCCTTCGAGCTTCAGTGAATTAATAC 999
 Oy 263 GluTyrGlyArgLeuLeuPheAlaAspProValValAsnGlnSerThrGlyIleThr 282
 Db 1000 AAACCTGGCGCAAAACTCTTTGGG----- 1023
 Oy 283 LeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetProGlyLeuTyrValArgVal 302
 Db 1024 -----CTTCCTCTGAGAAATGTGTTCTTGCA 1050
 Oy 303 LeuMetAspGlnVal---AlaValAspAsnAlaPheValValProGlnGlnAlaValThr 321
 Db 1051 ACACCTTGATGATTCAACAAAGTTTGATGCTGCTCGAAGTACGAGAGAACAGCTTATG 1110
 Oy 322 ArgGlyAlaLysAsp---ThrValMetIleValAsnAlaGlnGlyGlyMetGluProArg 340
 Db 1111 AAAACAACCTTCAGATGTACTACTGAAATGTCGATCACAAGATGTGAGCAAG----- 1164
 Oy 341 GluValThrValAlaGlnGlnGlnGlyThrAsnTrpIleValThrSerGlyLeuLysAsp 360
 Db 1165 -----ACAGCACCAAGCAACAAC----- 1182
 Oy 361 GlyAspLysValValValGluGlyLysSerIleAlaGlyIleThrGlyAlaLysVal 380
 Db 1183 -----GAAGGGCTTACGCGACAGCAGG-----CATGCT 1209
 Oy 381 ThrProLysGluTrpAlaSerSerGlu---AsnGlnAlaAlaAlaProGlnSerGlyVal 399
 Db 1210 TTGGAAACGGCGTGACTGTTTTCAGAGATTAGACAACTCTTAACCCCAAGATGCCATG 1269
 Oy 400 GlnThrAlaSerGlnAlaLysThrAlaSerGlu 410
 Db 1270 AAGCATCATCATCGCTTAACCGCTGATGAG 1302

RESULT 26
 US-10-156-761-1
 ; Sequence 1, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156, 761
 ; PRIOR FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 1
 ; LENGTH: 9025608
 ; TYPE: DNA
 ; ORGANISM: Streptomyces avermitilis
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (4187715)
 ; OTHER INFORMATION: a, t, c, g, other or unknown
 US-10-156-761-1

Alignment Scores:		53.3	Length:	9025608
Pred. No.:	128.50	Matches:	110	
Score:	32.92%	Conservative:	51	
Percent Similarity:	32.92%	Mismatches:	175	
Best Local Similarity:	6.36%	Indels:	154	
Query Match:	6.36%	Gaps:	19	
DB:	14			

US-09-889-756A-2 (1-412) x US-10-156-761-1 (1-9025608)

QY	7	LysAlaMetArg-----AlaAlaLeuAlaAlaAlaVal-----	18
DB	5815717	AAGATCATCTCGGTGCTCCCGGAACTGGGCGGCTGGCCAGAGCGGTCCGGGGGTG	5815776
QY	19	-----AlaLeuVal 21	
DB	5815777	AGCGCCCGGGGCGGGAGCCCGCTGAGCCCGCCGAGACCAAGCGGCTCTC	5815836
QY	22	LeuSerSerCysGly-----LysGlyValAspAlaLa 32	
DB	5815837	GTCCGGTGTGTGGCCGCGCACTCGGCTCGGCTACAGAGCCGCGCGGGGT	5815896
QY	33	GlnGlyGlnProAlaGly---ArgGlnAlaProAlaProValGlyValThr 51	
DB	5815897	CATGAGAGTCACCGGTGGGGGTGTGTGTACGCCAGCGCGGTGTACTCTGTGCTG	5815956
QY	52	Val-----HisProGlnThrValAlaLeuThrValGlu 62	
DB	5815957	GTCCGACACTGGTGTGCACACTGTGTGCACAGAGAACGCCAGACGCGTCTCGGGT	5816016
QY	63	LeuProGlyArgLeuGlnSerLeuArgThrAlaAspValArg-----	76
DB	5816017	GGAACGGGTCTGCTCCGTCTCGTCAAGTACTGAGAGAGACTTCAACCCCGCGCGGTCTCGA	5816076
QY	77	-----AlaGlnValGlyGlyTleGlnLysArgLeuPheGlnGly-----	91
DB	5816077	CCCGGTGAGACGCTGAGAGAACTTTGCCAGCGGTGCAAGTACAGAGGTGAGTCTGT	5816136
QY	92	-----SerTyrValAlaGlyGlnProLeu 100	
DB	5816137	GATGACGCCCCCGGAGACAGCGTCCGAGGTGCTCCCGGAACTGGGCGCGCG--	5816194
QY	101	TyrGlnIleAspSerSerThrTyrGlnAlaSerLeuGlnSerAlaArgAlaGlnLeuAla 120	
DB	5816195	GAAACGATGACGCCACACACACCGCGCTCGTGGCCACAGACCGAAGCCGACTGTAC	5816254
QY	121	ThrAlaGlnAlaThrLeuAlaLysAlaAspAlaAspLeuAlaArgTyrLysProLeu--	139
DB	5816255	GCCGAAACAATCAACGATGATCAAGGCCCGCTTCGCACTCGCGGCGCGGAACTCCGG	5816314
QY	140	ValAlaAlaGlnAlaValSerArgGlnLnuIzrAspAlaAla-ValThrAlaLysArgSe 159	
DB	5816315	GAGGCGCGGAGCGCGTGGCGGACCTCTCGGGCGCGCGCGCGCGCGCGCTCCAGC	5816374
QY	159	ValaGlnAlaGlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAs 179	
DB	5816375	C-----CGGGAGAGAGCGCGCGCGCAAGCGGTGACAGGCGCGCTCTCTGT	5816419
QY	179	ArgSerArgIleThrAlaProIleSerGlyPheIleGlyGln-----	193
DB	5816420	TGCGGACCGGTCTCCGCTCCCGCGCGCGGTTCACAGCACTTCTTCAGCAGTTCG	5816479
QY	194	-----SerLysValSerGlnGlyThrLeuAsnAlaGlyLysPheThrValLeuAlaThr 212	
DB	5816480	TGGCAGCGGGCGGAGAAAGCGCGGATCTTCGCGGTGACCAAGTTCGGGTTCAGCAGCA	5816539
QY	212	IleArgGlnThrAsnProMetTyrValAsnValThrGlnSerAlaSerGlnValMet--	231
DB	5816540	CGTGGCGGCGCAACGCGTCTGCGCTCCCGAGCTGACAGCCCGCTCGACGAGACGCTCT	5816599
QY	232	-----LysLeuArgArgGlnIleAlaGlu-----	239

Accession	Gene	Protein	Length	Score	Percent Similarity	Best Local Similarity
Dp	5816600	GGAGGAGCCATTGCGCTTGGCGGAGACC	CGGCGCTGATCTCCTCGGCGTGTGAAC	58166165		
Qy	240	-GlyValLeuLeuAlaAlaAspGlyValIleAlaValGlyIleLeuSphaAspGlyTh	259			
Dp	5816660	CGGCGACCGCGTAAACCGCGGAGAGCGATTGGCGCGCTCCCGTCCGCGCGCGAG	5816714			
Qy	259	rValTyr-rProGluysGlyArgLeuLeuPheAlaAspPro-----ValVala	276			
Dp	5816715	-----CAGCGCGCGCGCGCGCGCGGAGACAGACACCGAGCGGCTTGGCGGAGCGCA	5816764			
Qy	276	uSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetPr	296			
Dp	5816765	GGGGAGAGGCTCAGGAGCGCGCTCCGAGCGCGGATC-----GACCGGAGATCTCC	5816813			
Qy	296	oGlyLeuTyrValArgValLeuMetAspGlnValAlaValAspAsnAlaPheValValPr	316			
Dp	5816813	-----	5816813			
Qy	316	oGlnGlnAlaValThrArgGlyAlaAspThrValMetIleValAsnAlaGlnGlyI	336			
Dp	5816814	-CATCGGCGCGCTGGCCACCGCGGAGTGAAGACGGCGGCTCTGTC-----	5816858			
Qy	336	yMetGluProArgGluValThrValaIaGlnGlnGlyIleThrAsnTrpIleValThre	356			
Dp	5816859	-----GAGTCCAGCTCTCCACGCGGAGCGGGGTGCTGAGAGCTCGGAGAGG	5816905			
Qy	356	rGlyLeuLysAspGlyAspLysVal-----ValValGlnGlyIleSerIleAlaGlyI	374			
Dp	5816906	AGGCGAGAGTGGCGGACCGCGGCGGCTTCTGACGCGCGTGAACCTCGGAGGCG	5816963			
Qy	374	eThrcGlyAlaLysLysValThrProLysGluTrpAlaSerSerGluAsnGlnAlaAlaI	394			
Dp	5816964	-----	5816974			
Qy	394	aProGlnSerGlyValGlnThrAla	402			
Dp	5816975	GGTCCAGCGTGGCGCGCACACCGCC	5816999			
RESULT 27						
US-09-988-067B-51						
Sequence 51, Application US/09988067B						
Publication No. US20030124141A1						
GENERAL INFORMATION:						
APPLICANT: Haas, Rainer						
APPLICANT: Kleantous, Harold						
APPLICANT: Tomb, Jean-Francois						
APPLICANT: Miller, Charles						
APPLICANT: Al-Garawi, Amal						
APPLICANT: Odenbreit, Stefan						
APPLICANT: Meyer, Thomas						
TITLE OF INVENTION: Helicobacter Polypeptides and						
FILE REFERENCE: 06132/040002						
CURRENT APPLICATION NUMBER: US/09/988,067B						
CURRENT FILING DATE: 2003-01-31						
PRIOR APPLICATION NUMBER: US 98/831,309						
PRIOR FILING DATE: 1997-04-01						
NUMBER OF SEQ ID NOS: 112						
SOFTWARE: FastSeq for Windows Version 4.0						
SEQ ID NO 51						
LENGTH: 1116						
TYPE: DNA						
ORGANISM: Helicobacter pylori						
FEATURE:						
NAME/KEY: CDS						
LOCATION: (90)...(1076)						
US-09-988-067B-51						

Query Match: 6.32% Indels: 67
 DB: 11 Gaps: 6
 US-09-889-756a-2 (1-412) x US-09-988-067B-51 (1-1116)

Qy 63 Leuproglyargleugiserleuargthralaaspvalargalaglvalglylle 82
 Db 201 TTGCAGAGGTTTGGAGGCC---AGAGATAACAGCGTAGCTCCAAAGTCCCTGGCCG 257
 Qy 83 lleglnysargleupheglnuglyserlyrvalargalaglvalglylle 102
 Db 258 ATTGAAAAGGTTTGTAAAGGCGATCATTAAGGCGGATTTGGTTTTC 317
 Qy 103 llespserseerthryglualaasnleugiserleuargalaga----- 117
 Db 318 ATTCTAGCCCTGAATTGAAGCCAAACTCCCTCAGCTGAAGCCGGCATTAAGCCGCT 377
 Qy 117 ----- 117
 Db 378 AAAGCGCTTAGCGATGAAGTCAAAAGAGCTCAAGAGACGAATTAATTCTGCGAGA 437
 Qy 118 -----GlnleuAlaThrAlaGlnAlaThrleuAla----- 127
 Db 438 GACGTTTGGCAAGCAGCCAAATCCCAAGCCACTTTAGCCAAAGACTTATTAAGCGGCT 497
 Qy 128 -----LysAlaAspAlaAspLeuAlaArg 135
 Db 498 CAAGATTGTATGATTAATGCGCGCTGCGAGCTTGCAAAAGCCGATGAAGCCTATGCGCT 557
 Qy 136 TyrLysProleuValAlaAlaGlnAlaValSerArgGlnGluYrAspAlaAlaVal--- 154
 Db 558 TATGAAAGCAGCTAAATACAGAGAGCGCGCTTCCAAAGTATTAATAGCTTTAGCG 617
 Qy 155 -----ThrAlaLysSerAlaGlnAlaValSerArgGlnGluYrAspAlaAlaVal 170
 Db 618 GCGGCGAGCTCTGAAGTGAATGCTCCCTAAGGCTAAAGAGAGCGGCTTTAGCGCA 677
 Qy 171 lleyserleuaglylleasnleuasnargserargilethrAlaProAlaSerGlyPhe 190
 Db 678 GTGAATGAAGTGAAGCTTATTAAGACGTCGAAGCGACGCCCAATGATGAGGAA 737
 Qy 191 lleyglnserlyserleuValSerGlnGluYrleuLeuAlaGlyAspThrThrValleu 210
 Db 738 GTGAGTACGTCCTTTAAGCGGTGCGAGCTTACCTTAAGGGTTTCTGTGTTTGA 797
 Qy 211 AlathrilearglnthrasnprometyrvalaasnvalthrGlnserAlaSerGluVal 230
 Db 798 ATGATGATTTAAAGATGATGTTGTTAAATCAAGCGTGCCTGAAGATATTGTAACGAG 857
 Qy 231 MetLysleuargarglnlealaglglly----- 240
 Db 858 TTTAAAGTGGTAAAGAAATTT---CAAGGCTATATCCCGCGCTTAAAAAAGCAGCAA 914
 Qy 241 -----LysleuLeuAlaAlaAspGlyValAlaValAlaGlyLysPheAsp 257
 Db 915 TTCAAGGTCGAATATTGAGCGTGATGGGGATTTTGGCACTTGAAAGCGACGATATAT 974
 Qy 258 GlyThrValYrProGlnLysGlyArgleuLeuPheAlaAspProVal 273
 Db 975 TCCAACACTTACGACATGAAAGCTATGATGAGGACCATTCCTTA 1022

RESULT 28
 US-09-793-306-145
 ; Sequence 145, Application US/09793306
 ; Patent No. US20020098200A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Campos-Neto, Antonio
 ; APPLICANT: Skelky, Yasir
 ; APPLICANT: Ovendale, Pamela
 ; APPLICANT: Jen, Shyan
 ; APPLICANT: Lodes, Michael
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy

TITLE OF INVENTION: of Tuberculosis
 FILE REFERENCE: 014058-008740US
 CURRENT APPLICATION NUMBER: US/09/793,306
 CURRENT FILING DATE: 2001-02-26
 PRIOR APPLICATION NUMBER: US 60/185,037
 PRIOR FILING DATE: 2000-02-25
 PRIOR APPLICATION NUMBER: US 60/223,828
 PRIOR FILING DATE: 2000-08-08
 NUMBER OF SEQ ID NOS: 164
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 145
 LENGTH: 1794
 TYPE: DNA
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Description of Artificial Sequence: mttc#3-His
 OTHER INFORMATION: (MTB57)
 NAME/KEY: CDS
 LOCATION: (1)..(1794)
 OTHER INFORMATION: mttc#3-His (MTB57)
 US-09-793-306-145

Alignment Scores:
 Pred. No.: 0.000923 Length: 1794
 Score: 125.50 Matches: 91
 Percent Similarity: 31.54% Conservative: 50
 Best Local Similarity: 20.36% Mismatches: 164
 Query Match: 6.22% Indels: 142
 DB: Gaps: 14

US-09-889-756a-2 (1-412) x US-09-793-306-145 (1-1794)

Qy 1 MetAlaPheYrAlaPheYrAlaMetArgAlaAlaAlaAlaAlaValAlaLeu 20
 Db 175 TTGGCGGTCAAGTCTTGGAGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 234
 Qy 21 ValLeuSerSerCysGlyLysGlyAspAlaAlaGlnGlyGlnProAlaGlyArg 40
 Db 235 TAT-----GGGGGTGTTGGCTGCTCGCGGCGCGCGCGCGCGCGCGCGCG 279
 Qy 41 GlnAlaProAlaProAlaValGly-----ValAlaThrValHis 53
 Db 280 TCGGCTCAGGCAAGCGGCGGTGGCGATGCTTTGAGCGCGCGCGCGCGCGCGCGCAT 339
 Qy 54 ProGlnThrValAlaLeuThrValGluLeuProGlyArgLeuGlnSerLeuArgThrAla 73
 Db 340 CCGATGCTGTTGGCGGCCAAC-----CGTAATGCG 369
 Qy 74 AspValArgAlaGlnValGlyGlyLleGlnLysArgleuPheGlnGlySerTyr 93
 Db 370 TTGTGCGAG-----TTGGTGTTCGAATGCTTTGGCGCAATGCG--- 411
 Qy 94 ValArgAlaGlnProleuYrGlnLleAspSerSerThrYrGlnAlaAsnLeuGln 113
 Db 411 ----- 411
 Qy 114 SerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrleuAlaLysAlaAspAlaAspLeu 133
 Db 411 ----- 411
 Qy 134 AlaArgTyrLysProleuValAlaAlaGlnAlaValSerArgGlnGluYrAspAlaAla 153
 Db 412 -----CCGGGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCAT 462
 Qy 154 ValThrAlaLysArgSerAlaGlnAlaGlyValLysAlaAlaGlnAlaAlaLysSer 173
 Db 463 GTGGCCGGATGCGGCGCTATCAAGCGGGGATCGCGCGCGCGCGCGCGCGCTCGCG 522
 Qy 174 AlaGlyLeuLeuLeuAsnArgSerArgLleThrAlaPro-----LleSerGlyPheLle 191
 Db 523 TGTCAATGTGTGCGAGCGCGCTTCCACAGCTGCGCATCGCGCGCGCGCGCGCAT 582
 Qy 192 GlyLysSerLysValSerGlnGlyThrleu-----LeuAsnAlaGlyAsp-ThrThrVa 209

```

Db      583 GGCCTCGGCACATCGCTCGGAACTGGCGCGGGAACCCGGTGTACTACATCTG 642
Qy      209 lLeuAlaThrIleArgInThr-----AspProMetTy 220
Db      643 GGCACGGGAATTCGGGCAACCCCAAGTAGTACGGAACTCCGGCAAGCCCATGTG 702
Qy      220 r-----ValAsnValThrGln---Se 226
Db      703 GGCACGGGAATGACGTCGCCAAGATTGGGACGGGAATATGGGACCAACCATCTC 762
Qy      226 rAlaSerGluValMetLysLeuArgArgGlnIleAlaGluGlyLysLeu-----Al 244
Db      763 GGCACGGGAATGACGTCGCCAAGATTGGGACGGGAATATGGGACCAACCATCTC 822
Qy      244 aAlaAspGlyValIleAlaValGlyIleLysPheAspAspGlyThrValTyProGluLys 264
Db      823 GGCACGGGAATGACGTCGCCAAGATTGGGACGGGAATATGGGACCAACCATCTC 847
Qy      264 sGlyArgLeuLeuPheAlaAspProValAlaAsnGluSerThrGlyGlnIleThrLeuAr 284
Db      848 -----ACCTGGGCAAGTGGAAACCCGGAAGTACCAACTTC 882
Qy      284 gAlaAlaValProAsnAspGlnAsnIleLeuMetProGlyLeuTyValArgValLeuMe 304
Db      883 GCGCGCG----- 889
Qy      304 tAspGlnValAlaValAlaAspAsnAlaPheValValProGlnGlnAlaValThrArgGlyAl 324
Db      890 -----GAAATCTCGGTTCTTCAACTGGGCAAGTGGAAACATCGCTCTCCACATCGGT 945
Qy      324 aLysAspThrValMetIleValAsnAlaGlnGlyGlyMetGluProArgGluValThrVa 344
Db      946 TTCGGAACAACGCGGACATTAACCTCGCTCGGACCAATGGCAACAACAATCGGT 1005
Qy      344 lAlaGlnGlnGlnGlyThrAsnTrpIleValThrSerGlyLeuLysAspGlyAspLysVa 364
Db      1006 TTTCGGCTCACCGGCAACATCTGGTGGCATTTGGCGGTGAACCTCGGCATCGGAT 1065
Qy      364 lValValGluGlyIleSerIleAlaGlyIleThrGlyAlaLysValThrProLysG 384
Db      1066 CTAGGTTTCGGGAACCTCGGTAACAACAACATCGGTTCTTCAACTCGGCAACAACAAC 1125
Qy      384 uTrpAlaSerSerGluAsnGlnAlaAlaProGlnSerGlyValGlnThrAlaSerG 404
Db      1126 GTGGGCTTCTTCAATTCG-----GCAACAACAATTCGCTTGGAAACGCGGCGAC 1179
Qy      404 uAlaLysThrAlaSerGlu 410
Db      1180 ATCAACACGCGCTTCGGA 1198

```

RESULT 29

```

US-09-847-513A-1
; Sequence 1, Application US/09847513A
; Publication No. US20030104375A1
; GENERAL INFORMATION:
; APPLICANT: MBARI
; APPLICANT: DeLong, Edward
; APPLICANT: Beja, Oded
; TITLE OF INVENTION: Light-driven energy generation using proteorhodopsin
; FILE REFERENCE: MEA-101
; CURRENT APPLICATION NUMBER: US/09/847,513A
; PRIOR FILING DATE: 2001-05-01
; PRIOR APPLICATION NUMBER: 60/201,602
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: Patent version 3.0
; SEQ ID NO 1
; LENGTH: 105184
; TYPE: DNA
; ORGANISM: Naturally occurring gamma proteobacterium
; FEATURE:
; NAME/KEY: gene

```

```

; LOCATION: (50866)..(51615)
; OTHER INFORMATION: Proteorhodopsin gene sequence.
; NAME/KEY: misc feature
; LOCATION: (1593)..(2807)
; OTHER INFORMATION: Predicted threonine dehydratase. Contains 'n' at position 2753.
; PUBLICATION INFORMATION:
; AUTHORS: Beja, O., Aravind, L., Koonin, E.V., Suzuki, M.T., Hadd, A., Nguyen, L.P.,
; AUTHORS: Jovanovich, S.B., Gates, C.M., Feldman, R.A., DeLong, E.F.
; TITLE: Bacterial rhodopsin: evidence for a new type of phototrophy in the sea
; JOURNAL: Science
; VOLUME: 289
; ISSUE: 5486
; PAGES: 1902-1906
; DATE: 2000-09-15
; DATABASE ACCESSION NUMBER: AF279106
; DATABASE ENTRY DATE: 2000-06-15
; RELEVANT RESIDUES: (50866)..(51615)
US-09-847-513A-1

```

Alignment Scores:

Pred. No.:	0.281	Length:	105184
Score:	125.00	Matches:	72
Percent Similarity:	38.01%	Conservative:	50
Best Local Similarity:	22.43%	Mismatches:	147
Query Match:	6.19%	Indels:	52
DB:	11	Gaps:	9

US-09-889-756a-2 (1-412) x US-09-847-513A-1 (1-105184)

```

Qy      58 AlaLeuThrValGluLeuProGlyArgLeuGluSerLeuArgThrAlaAspValArgAla 77
Db      88348 GCGTAACTAGAGAAATTCCTGGAAGCTCATTCACCAACGCTAAGTACATTT 88407
Qy      78 GluValGlyIleIleGlnLysArgLeuPheGlnGluGlySerTyValArgAlaGly 97
Db      88408 GAAATACCTCGAAGATAACTAATTAATGATATCGAGATGAGTCACTTTAGG 88467
Qy      98 GluProLeuTyrglnIleAspSerSerThrTyGluAlaAsnLeuGluSerAlaArgAla 117
Db      88468 GATGAATCTCCCTCATTTGATGAT-----ACAGAGCTTTAGCA 88506
Qy      118 GluLeuAlaThrAlaGlnAlaThrLeuAlaLysAlaAspAlaAspLeuAlaArgTyLys 137
Db      88507 CAATTAATCAATCAACAAGCTAATTTGATTAGCTGAACAACATCTAGCAAGATATATA 88566
Qy      138 ProLeuValAlaAlaGluAlaValSerArgGlnGluTyTrpAlaAlaValThrAlaLys 157
Db      88567 AATCTCAGACAGATGCGCATATTCTATTCAAGATCTTGAT----- 88608
Qy      158 ArgSerAlaGluAlaGlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsn 177
Db      88609 ---AAGCTGAATCTGATCAATAGTAGTACGTCGACGATGATTTTATAGAGTTAA 88665
Qy      178 LeuAsnArgSerArgIleThrAlaProIleSerGlyPheIleGlyGlnSerLysValSer 197
Db      88666 TTGGACAAACTAGTACTAGCTCTTTAATGAGGATTATCAAAATGATTCTAGAT 88725
Qy      198 GluGlyThrLeuLeuAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAsn 217
Db      88726 ACAGATCAAGTATCAATGACGCT---GTCCAAGTTTGAATTTTAGGCTCTACCAAT 88782
Qy      218 -----Promet---TyValAsnValThrGlnSerAlaSerGlu 229
Db      88783 GTTGAAGCAAGATTTCTATTCTCAATGAACATATACAGGTTAAGTTGAGATGAG 88842
Qy      230 ValMetLysLeuArgGlnIleAlaGluGlyLysLeuLeuAlaAlaAspGlyValIle 249
Db      88842 ----- 88842
Qy      250 AlaValGlyIleLysPheAspAspGlyThrValTyProGluLysGlyArgLeuLeuPhe 269
Db      88843 -----TATCAATTTGATATCAGAGAGATA---TCTACAAAGCTACGTTAGAGAGA 88890

```

```

Qy 270 AlaaspProValValAsnGluSerThrGlyGlnIleThrLeuArgAlaValProAsn 289
Db 88891 CTGGCTCCCATG-----TCAACCGAGGCTCCCAATATAGGTATGCAATTTTGA 88941
Qy 290 AspGlnAsnIleLeuMetProGlyLeuThrValArgValLeuMetAspGlnValAlaVal 309
Db 88942 TTTGATACCTTTTAAATCCAGATCAATAGTAAAGCTCAATCAAGCATCACTGAGAA 89001
Qy 310 AspaenAlaPheValAlaProGlnIleAlaValThrArgGlyAlaLeuAsp-----Thr 327
Db 89002 GCAAAAGGTACATGGTTCATTAAGTCACTGCTCCAGTCCGAAACAAGTATTTGGGCC 89061
Qy 328 ValMetIleValAsnAlaGlnGlyGlyMetGluProArgGluValThrValAlaGlnGln 347
Db 89062 ATATATACCATTAACGACCAACAGTATGTTAGAGATCTGTGATGTTATTATTTT 89121
Qy 348 GlnGlyThrThrThrIleValIleThrSerGlyLeuLeuAspGlyAspValValAlaGlu 367
Db 89122 GAAGAGCAATATGCTTTTGTACGCGAACACTATATATGATTAATTTTATAGGC 89181
Qy 368 G1Y 368
Db 89182 GGA 89184

RESULT 30
US-08-945-038-3
Sequence 3, Application US/08945038
Publication No. US20020146423A1
GENERAL INFORMATION:
APPLICANT: Doidge, Christopher V.
APPLICANT: Lee, Adrian
APPLICANT: Radcliffe, Fiona J.
APPLICANT: Hocking, Diana M.
APPLICANT: Webb, Elizabeth A.
TITLE OF INVENTION: PROTECTIVE HELICOBLACTER ANTIGENS
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY & LARDNER
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/945,038
FILING DATE: 23-DEC-1997
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/AU96/00225
FILING DATE: 19-APR-1996
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: AU PN2575
FILING DATE: 21-APR-1995
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: AU PN3931
FILING DATE: 03-JUL-1995
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: AU PN7565
FILING DATE: 16-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Bent, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 017227/0133
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 990 base pairs

```

```

/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/ ORIGINAL SOURCE:
/ ORGANISM: helicobacter pylori
/ IMMEDIATE SOURCE:
/ CLONE: clone E2.5
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 1..987
/ US-08-945-038-3

Alignment Scores:
Pred. No.: 0.000859 Length: 990
Score: 122.50 Matches: 63
Percent Similarity: 36.23% Conservative: 37
Best Local Similarity: 22.83% Mismatches: 109
Query Match: 6.07% Indels: 67
Gaps: 6

US-09-889-756a-2 (1-412) x US-08-945-038-3 (1-990)
Qy 63 LeuProGlyArgLeuGluSerLeuArgThrAlaAspValArgAlaGlnValGlyGlyIle 82
Db 112 TTYGCAAGGATTTTGTGAAGCC--AGAAATACAGCGTGATTCCAAAGTCCCTGGCCG 168
Qy 83 IleGlnIleValArgLeuPheGlnGlySerThrValArgAlaGlnIleProLeuThrGln 102
Db 169 ATTGAAGAAGGTGTTGTTAAAGGCGATCGCATTAAGGCGCATTTGGTTTAC 228
Qy 103 IleAspSerThrThrGlyAlaAsnLeuGluSerAlaArgAla----- 117
Db 229 ATTCTAGCCCTGATATAGAACCAAGCTCGTCAAGCTGAAGCGGCAATTAAGCGCT 288
Qy 117 ----- 117
Db 289 AAAGCGTTAGCGATGAGTCAAAAGAGCTCAAGAGCAAGCAATTCAGAGA 348
Qy 118 -----GlnLeuAlaThrAlaGlnAlaThrLeuAla----- 127
Db 349 GACGTTGGCAAGCGCAAAATCTCAAGCCATTACCAAGAGACTTATAGCGGCTT 408
Qy 128 -----LysAlaAspAlaAspLeuAlaArg 135
Db 409 CAAGTTTGATATATATGCGGTGCGAGCTTGCAAAAGGCGCATGAAGCTATAGCGGCT 468
Qy 136 TyrIlePheLeuValAlaAlaGluAlaValSerArgGlnIleThrAspAlaAlaVal--- 154
Db 469 TATGAAAGCACTAAATCAACAGAGCGCGCTTACCAAAAGTATTAATGCGCTTTAGCG 528
Qy 155 -----ThrAlaLysArgSerAlaGluAlaGlyValLysValAlaAlaAla 170
Db 529 GGGCGGAGCTCTGAAGATTAAGTTCGCTAAGCTTAAGAGAGCGGCGCTTTAGGCGCA 588
Qy 171 IleLysSerAlaGlyIleAsnLeuAsnArgSerArgIleThrAlaProIleSerGlyPhe 190
Db 589 GTGAATGAAGTGAAGTCTTATTTAAAGATGTCAAGACGACGCCAATGATGAGGGA 648
Qy 191 IleGlyIleSerLysValSerGlyGlyThrLeuLeuAsnAlaGlyAspThrThrValLeu 210
Db 649 GTGAGTATGTGCTTTAAGCGGTGCGAGCTTACCCCTTAAGGCGCTTCTGTGTGCTC 708
Qy 211 AlaThrIleArgGlnThrAsnProMetThrValAsnValThrGlnSerAlaSerGluVal 230
Db 709 ATGATGATTTAAAGATAGTGTGTTAAATAACGCGCTGAAGATTTGAACGAT 768
Qy 231 MetLysLeuArgArgGlnIleAlaGluGly----- 240
Db 769 TTTAAAGTGGTAAAGAAATT--GAAGTTATATCCCGCGTTGAAGAAAGCGCGAAA 825
Qy 241 -----LysLeuLeuAlaAlaAspGlyValIleAlaValGlyIleLysPheAsp 257

```

Db 826 TTCAGGCTCAATATTGACGCTGATGGGGGATTTTGGCACTTGGAAGCAGCAATAT 885
 Qy 258 GlyThrValTyrProGluLeuGlyArgLeuLeuPheAlaAspProVal 273
 Db 886 TCACAACACTTACGACATGATAAAGCATGATGAGGAGCGCATACCTTA 933
 RESULT 31
 US-10-156-761-413
 / Sequence 413, Application US/10156761
 / Publication No. US20030119018A1
 / GENERAL INFORMATION:
 / APPLICANT: OMURA, SATOSHI
 / APPLICANT: IKEDA, HARUO
 / APPLICANT: ISHIKAWA, JUN
 / APPLICANT: HORIKAWA, HIROSHI
 / APPLICANT: SHIBA, TADAYOSHI
 / APPLICANT: SAKAKI, YOSHIYUKI
 / APPLICANT: HATTORI, MASAHIRA
 / TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 / FILE REFERENCE: 249-262
 / CURRENT APPLICATION NUMBER: US/10/156,761
 / CURRENT FILING DATE: 2002-05-29
 / PRIOR APPLICATION NUMBER: JP 2001-204089
 / PRIOR FILING DATE: 2001-05-30
 / PRIOR APPLICATION NUMBER: JP 2001-272697
 / PRIOR FILING DATE: 2001-08-02
 / NUMBER OF SEQ ID NOS: 15109
 / SEQ ID NO 413
 / LENGTH: 5505
 / TYPE: DNA
 / ORGANISM: Streptomyces avermitilis
 / FEATURE:
 / NAME/KEY: CDS
 / LOCATION: (1)..(5505)
 US-10-156-761-413
 Alignment Scores:
 Pred. No.: 0.00907 Length: 5505
 Score: 122.50 Matches: 120
 Percent Similarity: 33.14% Conservative: 50
 Best Local Similarity: 23.39% Mismatches: 174
 Query Match: 6.07% Indels: 170
 DB: 14 Gaps: 20
 US-09-889-756a-2 (1-412) x US-10-156-761-413 (1-5505)
 Qy 7 LysAlaMetArgAlaAlaAlaLeuAlaAlaValAlaLeuValLeuSerSerCysGly 26
 Db 1083 AGAGCGGGAACGCGCGCTGCTCGCTCGCTCGCAAGTCCAAACATCGATCACAGCAGGC 1142
 Qy 27 LysGly-----GlyAspAlaAlaGln----- 33
 Db 1143 CGCGCGGGGTGTCCCGGTGTGATCAAGATGTCGTGCGATGACGAGCAGCGGTGTGCC 1202
 Qy 33 ----- 33
 Db 1203 GAAGACCTTCACGTGACCGCGCGGCGGTGACGTGTCGTGCGCTCGCTCGTCA 1262
 Qy 34 -----GlyGlyGlnProAlaGlyArgGluAlaProAlaProValGlyVal 49
 Db 1263 CTGTGTCAGCGAGCGCATGCGGTGCGCGAGACGAGACACCTCGCGCGCGCGGTCTC 1322
 Qy 50 Val-----ThrValHisProGlnThrValAlaLeuThr--- 60
 Db 1323 GTCTTTCGGATCAGCGCGACCAATGCGCACATATCTTCAGAGCAGCGCGCGAGCCGGA 1382
 Qy 61 ValGluLeuProGlyArgLeuGluSerLeuArgThrAlaAspValArgAlaGlnVal--- 79
 Db 1383 GCGCGGGAGACCGCGCCCATGACACGCTGAAGGAGGAGCGGTGCGCAGGAACTGCTGCG 1442
 Qy 80 -----GlyGlyIleIleGlnLysArgLeuPheGlnGlySerTyrValArg 95
 Db 1443 GCGCCGAGCACGAGGTGTGCGGTGTACTCTCCGCGAAGATTGGGCGCGCTGCGAGA 1502

Qy 96 AlaGly-GlnProLeuTyrGlnIleAspSerSerThrTyrGluAlaAsnLeuGluSerAl 115
 Db 1503 GCAAGCACAACGACTGCTG-----TCGTCGGAGCACTCGC 1538
 Qy 115 ArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaLeuAlaAspAlaAspLeuAlaAr 135
 Db 1539 GCGCGCTTCACCGGCGCCACGTCGCGGTTCGCTGCGCGCCACCGCTCGCGGTGAGCA 1598
 Qy 135 gTyrLysProLeuValAlaAlaGluAlaValaLaserArgGlnGluTyrAspAlaAlaValTh 155
 Db 1599 CCGGCGCGGTGTCTCGCGCGCAGCGG-----GAAGAGCTCGCGCGATCTTGAC 1649
 Qy 155 LAlaLysArgSerAlaGluAlaGlyValLysAlaAlaGlnAlaAlaLysSerAlaG 175
 Db 1650 CGCC-----CTCGCGCGCGCGGAGCCCGCGCAGAGTGTACAGCGCGCGCGCG 1700
 Qy 175 YLLeuLeuAsnArgSerArgIleThrAlaProIleSerGlyPheIleGlyGlnSerLys 195
 Db 1701 GCGCGACCG-GGCGAAGTCGGGTCTCTGTCGCGGTACAGGTTTC----- 1745
 Qy 195 sValSerGlnGlyThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIleArg 215
 Db 1746 -----GAGCGGATCGCATGGGCGCGGACTTGACGA-- 1778
 Qy 215 nThrAsnProMetTyrValAsnValThrGlnSerAlaSerGluValMetLysLeuArgAr 235
 Db 1779 -----GGCTACCCCGTTCGCGC 1798
 Qy 235 gGlnIleAlaGlnGlyLysLeu-----Le 243
 Db 1799 CGGCTTCACAACGAGTGTGCGCGCTGTGACTTGCCTGCGATGTCGATCGGAACGC 1858
 Qy 243 uAlaAlaAspGlyValIleAlaValGlyIleLysPheAspAspGlyThrValTyrProG 263
 Db 1859 TTCATCAGACGGGTTCGACGCGACCGCGGTTCGCGATGAGAGTGCCTGTTCCGCT 1918
 Qy 263 uLysGlyArgLeuLeuPhe---Ala-AspProValAlaGlnSerThrGlyGlnIle- 281
 Db 1919 TGCTGAGTGTGTGGCCATTCGCGCGGATTAAGTGGCGGTACTCGGTGGAGAGATTG 1978
 Qy 282 -----ThrLeuAla 284
 Db 1979 CTGCGCGCAGCTGCGCGGGGTGTGTCGCTCGGAGATGCGGAGAGCTGTGTCGCTGC 2038
 Qy 284 rGAlaAlaValProAlaAspGlnAsnIleLeuMetProGlyLeuTyrValArgValLeu 304
 Db 2039 GTGCGCGGTGATG-----CAGGCGTTGCCCGCGCGCGCATGTGTCGCGT- 2088
 Qy 304 eAspGlnValAlaValAspAsn-----A 312
 Db 2089 -----CAGGCGCGCGAGAGAGAGTGTGCTGCTGCGGATGCGGAGAGTGGCATCGCGG 2143
 Qy 312 lApheValValProGlnGlnAlaValThrArgGlyAlaLysAspThrValMetIleValAla 332
 Db 2144 CGATCAACGGCCTTCAGTGTGTGTGTCGCGCGCGCAGAGCAGGAGTATGGCGTT- 2202
 Qy 332 snAlaGlnGlyMetGluProArgGluValThrValAlaGlnGlnGlnGlyThrAsnT 352
 Db 2203 -----GCCGAGGTGTTCTGCTCAGGCGCGTAAACCTCC 2239
 Qy 352 rIleValThrSer-----GlyLeuLysAspGlyAspL 363
 Db 2240 GATTGAAGTCATTCACGCGTTTCATTCCCGCTTGATGATCCGATGTTGGAGAGATTGC 2299
 Qy 363 yValValValGlnGlyIleSer-----IleAlaGlyIleT 375
 Db 2300 CGCGGTTGTCGCGGTTCAGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 2359
 Qy 375 hGlyIly-----AlaLysLysValThrProLysGluThrAlaSerSerGluAsnGlnAla 393
 Db 2360 CGGCGCGCTGCGCGAGCGGTACACCCCGAGACATGAGTTCGCGCAGTTCGTCGAGCGG 2419

; PRIOR FILING DATE: 1999-08-08
 ; NUMBER OF SEQ ID NOS: 68
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 1
 ; LENGTH: 12145
 ; TYPE: DNA
 ; ORGANISM: Klebsiella pneumoniae
 US-10-277-249-1

Alignment Scores:

Pred. No.:	0.0345	Length:	12145
Score:	121.50	Matches:	80
Percent Similarity:	37.28%	Conservative:	49
Best Local Similarity:	23.12%	Mismatches:	136
Query Match:	6.02%	Indels:	81
DB:	12	Gaps:	17

US-09-889-756a-2 (1-412) x US-10-277-249-1 (1-12145)

```

Qy 69 SerleuArgThrAlaAspValArgAlaGlnValGlyGlyIleIleGlnLysArgLeuPhe 88
Db 9662 ACAGTGATGCCGCTTTGTCCGGAGTCGGCGGAAGTATCAGACGGCATAGCTG 9721
Qy 89 GlnGluGlySerTyrValArgAlaGlyGlnProLeuTyrGlnIleAspSerSerThrTyr 108
Db 9722 CGTAAAGAGACTAAGCGAGGTCAGCATGCCGTTAATA----- 9760
Qy 109 GluAlaAsnLeuGluSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaLys 128
Db 9761 ---GCCGGATTGATATCGGCAC-----GCCACCACCGAGTGGCGCTGCG--- 9805
Qy 129 AlaAspAlaAspLeuAlaArgTyrLysProLeuValAlaAlaGlnAlaValSerArgGln 148
Db 9806 -----TCCGACTACCCCGACGCGGCGGCGTTTGTTCACGCGGATGTCGCG----- 9853
Qy 149 GluTyrAspAlaAlaValThrAlaLysArgSerAlaGlnAlaGlyValLysAlaAla--- 167
Db 9854 -----ACGACGGCGCATGAAGGAGCGGGGACATATTCGCGGAGCCCTGCGCGCTG 9907
Qy 168 ---GlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsnArgSerArgIle----- 183
Db 9908 GAGCAGCGCTCGCGCAAAACACCGTGTGATGACGATGCTCTCGCATCTATCTTAAC 9967
Qy 184 ---ThrAlaProIleSerGlyPheIleGlyGlnSerLysValSerGluGlyThrLeuLeu 202
Db 9968 GAAGCGCGCGCGTATGCGCATGCGCATGCGACCAATCACCAG--ACCTTATTC 10024
Qy 203 AsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAsnProMetTyrValAsn 222
Db *10025 ACC-----GAATCGACCATGTCGTCATTAACCGCAGACG---CCGGCGCGGCTGGGC 10075
Qy 223 ValThrGlnSerAlaSerGluValMetLysLeuArgGly----- 235
Db 10076 GTT-----GGCGTGGGAGACGATATCGCCCTCGCGCGCTGCGACGCTGCGCGCGCG 10129
Qy 236 GlnIleAlaGlnGlyLysLeuLeuAlaAlaAsp----- 246
Db 10130 CAGTATGCCAGAGGGGTGATCTACTGATGACGACCGCGCTGATTTCTTGAGCCGCTG 10189
Qy 247 -----GlyValIleAlaValGlyIleLysPhe 255
Db 10190 TGTGTGCTCAATGAGCGCTGACCGGGGATCAACGATGCGCGCGCATCTCTCAAAAG 10249
Qy 256 AspAspGlyThrValTyrProGlnLysGlyLeuLeuPheAlaAspProValValAsn 275
Db 10250 GACGCGCGCTGCTGGTG-----AACAAACCGCTCGCTTAAACCTTCGTTGTGTGAT 10303
Qy 276 GluSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAsp-----Gln 291
Db 10304 GAA-----GTACGCTGCTGAGACAGGTCGCCGAGGGGCTATGCGCGCGGTG 10351
Qy 292 AsnIleLeuMetProGlyLeuTyrValArgValLeuMetAspGlnValAlaAspAsn 311
  
```

```

Db 10352 GAATGCGCGCGCGCGAGTGTGCGATCTCTGCAATCCCTACGGGATGCCACC 10411
Qy 312 AlaPhe-----ValValProGlnGlnAlaValThrArg 322
Db 10412 TTCTTCGGGCTTAACCCCGAAGAACCCAGGCCATGCTCCCATCGCCCGCCTGATT 10471
Qy 323 GlyAlaLysAspThrValMetIleValAlaAlaGlnGlyMetGluProArgGluVal 342
Db 10472 GGCAACGTTCCGGGTGCTCAAGACCCCGAGGGGATGTGACGTGCGGGTGATC 10531
Qy 343 ThrValAlaGlnGlnGlyThrAsnThrPheValThrSerGlyLeuLysAspGlyAsp 362
Db 10532 CCGCGCGCGC-----AACCTTCACTTAAGCGCGCAAAAGCCGCGAGAG 10576
Qy 363 LysValValValGluGly 368
Db 10577 GCCGATGTCCCGGAGGCG 10594

RESULT 35
US-10-213-203-1
; Sequence 1, Application US/10213203
; Publication No. US200300827556A1
; GENERAL INFORMATION:
; APPLICANT: BURCH, ROBERT R.
; APPLICANT: DORSCH, ROBERT R.
; APPLICANT: LAFFEND, LISA ANNE
; APPLICANT: NAGARAJAN, VASANTHA
; APPLICANT: NAKAMURA, CHARLES
; TITLE OF INVENTION: 1,3-PROPANEDIOL AND POLYMER DERIVATIVES FROM A
; FILE REFERENCE: CR-9715-F
; CURRENT APPLICATION NUMBER: US/10/213,203
; PRIOR FILING DATE: 2002-08-05
; CURRENT APPLICATION NUMBER: 08/440,293
; PRIOR FILING DATE: May 12, 1995
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 1
; LENGTH: 12145
; ORGANISM: Klebsiella pneumoniae
US-10-213-203-1

Alignment Scores:
Pred. No.: 0.0345 Length: 12145
Score: 121.50 Matches: 80
Percent Similarity: 37.28% Conservative: 49
Best Local Similarity: 23.12% Mismatches: 136
Query Match: 6.02% Indels: 81
DB: 14 Gaps: 17

US-09-889-756a-2 (1-412) x US-10-213-203-1 (1-12145)
Qy 69 SerleuArgThrAlaAspValArgAlaGlnValGlyGlyIleIleGlnLysArgLeuPhe 88
Db 9662 ACAGTGATGCCGCTTTGTCCGGAGTCGGCGGAAGTATCAGACGGCATAGCTG 9721
Qy 89 GlnGluGlySerTyrValArgAlaGlyGlnProLeuTyrGlnIleAspSerSerThrTyr 108
Db 9722 CGTAAAGAGACTAAGCGAGGTCAGCATGCCGTTAATA----- 9760
Qy 109 GluAlaAsnLeuGluSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaLys 128
Db 9761 ---GCCGGATTGATATCGGCAC-----GCCACCACCGAGTGGCGCTGCG--- 9805
Qy 129 AlaAspAlaAspLeuAlaArgTyrLysProLeuValAlaAlaGlnAlaValSerArgGln 148
Db 9806 -----TCCGACTACCCCGACGCGGCGGCGTTTGTTCACGCGGATGTCGCG----- 9853
Qy 149 GluTyrAspAlaAlaValThrAlaLysArgSerAlaGlnAlaGlyValLysAlaAla--- 167
Db 9854 -----ACGACGGCGCATGAAGGAGCGGGGACATATTCGCGGAGCCCTGCGCGCTG 9907
  
```


Db 3395 CTTGGCGTAAAGCGCTGTACGTTTGTCTGAG 3424

RESULT 38
US-10-156-761-1788
; Sequence 1788, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1788
; LENGTH: 2088
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2088)
US-10-156-761-1788

Alignment Scores:
Pred. No.: 0.00505 Length: 2088
Score: 119.50 Matches: 97
Percent Similarity: 36.28% Conservative: 63
Best Local Similarity: 22.00% Mismatches: 162
Query Match: 5.92% Indels: 119
DB: 14 Gaps: 20

US-09-889-756a-2 (1-412) x US-10-156-761-1788 (1-2088)

Qy 52 ValHisProGlnThrValAlaLeuThrValAlaLeuPro----- 64
:::|||||
481 CTGCATCCG-----GCCCGCTCGTCGTCAGTCCGATCGGCACCCAGACGGGTTCC 534
65 -----GlyArgLeuGlnSerLeuArgThrAlaAspValArgAlaGln 78
535 AGGGAGTGTGATCTGTGTCGATGAGTCTGCTGCTGCGGCGGACCGGACACC 594
Qy 79 Val---GlyGlyIleIleGlnLeuArgLeuPheGlnGlnGlySerTyrValArgAlaGly 97
|||::|
595 GTCCAGAGAGGGCCGGTCCGCGACCCCTGCGGAGCAGACGGAC-----CGCGCCGA 648
Qy 98 GlnProLeuTyrGlnIleAspSerSerThrTyrGlnAlaAsnLeuGlnSerAlaArgAla 117
:::|
649 CGCGAGCTGAGAGAGCGGTGCGGAACCTCATCCCGTCCGCGTGAAGAGTTCGGCG 708
Qy 118 GlnLeuAlaThrAlaGlnAlaThrLeuAlaLysAlaAspAlaAspLeuAlaArg----- 135
:::|
709 CAGGCGACGCTCTCCGACAGACCTTCACTCCGCGGTCCGACCTGACCCGTACCGGG 768
Qy 136 -----TyrLysProLeuVal 140
Db 769 GACGCGTGTCTGCTCTGCGCTCCGCTACCGCAATCCGCGCATCGAACCCGCTGCTG 828
Qy 141 AlaAlaGlnAlaValSerArgGlnGlnTyrAspAlaAlaValThrAlaLysArgSerAla 160
829 GCTGCC---GCCGTGGCTTACCTGCTCCGCGGTGAGCAGTACCGGCGTACCGGACCC 885
Qy 161 GlnAlaGlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsnArg 180
886 GAGAGCGCGCGGTGCGGAGACGGGCGCGCCGACCCGCGGCGCGCTTCGCGGCCCTGCTG 945

Qy 181 SerArgIleThrAlaProIleSerGlyPheIleGlnSerLysValSerGlnGlyThr 200
:::|
Db 946 TTCAAGTGAACCGGACCGGACGAGGCGGCTGACGTATCGCTGTACTCCGGAACG 1005
Qy 201 LeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIleArgIleThrAsnProMetTyr 220
:::|
Db 1006 ---ATCGGAAAGAGAGACCGTCTGACACCGCGCGCGGACGAGACCGGATCGGC 1062
Qy 221 ValAsnValThrGlnSerAlaSerGlnValMetLysLeuAlaGArgGlnIleAlaGly 240
:::|
Db 1063 CGGATTCGCTGCTCCAGCCGACCCGATCCACACGAGACTGACCGGCGGCTGCC---GGG 1119
Qy 241 LysLeuLeuAla----- 244
Db 1120 GACATCGTGCCGTCGTGCGGCGGAGACCGCGCGCGGCTGACCCCTGTGGCAACC 1179
Qy 245 -----AlaAspGlyValIleAlaValGlyIle 253
Db 1180 GCGGCTCCCTTGTCTCGAACCGCGACCGCTCCGATCCGTCGTCCTCGCGGCTC 1239
Qy 254 LysPheAspAspGlyThrValTyrProGlnLysGlyArgLeuPheAlaAspProVal 273
Db 1240 GAGCGCGCGACAGCACCC-----GACACCGACCGCTTGCGCTCGCGCGCGC 1290
Qy 274 ValAsnGlnSerThrGlnIleThrLeuArgAlaAlaValProAsnAspGlnAsnIle 293
Db 1291 CTGCGCCAGAGAGATCCCTCGCTGCTGTCGAGACCCGACCGGACCGGCTGACCTG 1350
Qy 294 Leu----MetProGlyLeuTyrValArgValLeuMetAspGlnValAlaValAspAsn 311
Db 1351 CTGTCGGGCAATGGGCGAGCTCATCTGAGGTGGCGGTGGAGAGATCCGCGGACCGG 1410
Qy 312 AlaPheValAlaPro-----GlnGlnAlaValThrArgGly 323
Db 1411 GGAACGCCCGTCAATGTGGCGCGGCGGAGGTCCTTACCGGAGACGTCGCCGCGGC 1470
Qy 324 AlaLysAspThrVal----- 328
Db 1471 GTGCCGGGCTCGTACCCGATGTCAAACAGACGCGCGCGCGCCACTTGGCCAT 1530
Qy 329 MetIleValAsnAlaGln-----GlyGlyMetGlu 338
Db 1531 GTGCTCTCAGCGAGCGCTGAGTCCGCGCGGACGAGCGGCGGCGGCTTGTGAG 1590
Qy 339 ProArgGlnValThrValAlaGlnGlnGlnGlyThrAsnTrpIle-----ValThrSer 356
Db 1591 TTCGCTCGGCGCTGTCGCGGCGGCGTCCGACGAGTACGTCGCGCGGTGCGGCC 1650
Qy 357 GlnLeuAspGlyAspLysValValIleGlnGlyIleSerIleAlaGly-----Ile 374
Db 1651 GGTCTCGGGAGC-----GCCCTGCGGAGAGG---CCGCTCGCGGTACCCCGGTG 1698
Qy 375 ThrGlyAlaLysLysValThr-----ProLysGlnTrpAlaSerSer 388
Db 1699 ACCGGGCTGAGAGTCTGTCGACCGATGAGCAGCACCCATCCGAAGAC-----TCTTCC 1752
Qy 389 Glu 389
Db 1753 GAG 1755

RESULT 39
US-10-190-435-51/C
; Sequence 51, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF

FILING DATE: 01-Apr-2002
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 09/453,702
 FILING DATE: 03-DEC-1999
 APPLICATION NUMBER: 60/110,955
 FILING DATE: 04-DEC-1998
 ATTORNEY/AGENT INFORMATION:
 NAME: Seay, Nicholas J.
 REGISTRATION NUMBER: 27386
 REFERENCE/DOCKET NUMBER: 960296.95017
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (608) 251-5000
 TELEFAX: (608) 251-9166
 INFORMATION FOR SEQ ID NO: 72:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 46819
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 SEQUENCE DESCRIPTION: SEQ ID NO: 72:
 US-10-114-170-72

Alignment Scores:
 Pred. No.: 0.362 Length: 46819
 Score: 119.50 Matches: 100
 Percent Similarity: 35.05% Conservative: 50
 Best Local Similarity: 23.36% Mismatches: 165
 Query Match: 5.92% Indels: 114
 Gaps: 20

US-09-889-756a-2 (1-412) x US-10-114-170-72 (1-46819)
 QY 11 AAlaAlaLeuAlaAlaValAlaLeuSerSerCysGlyGlyGlyAsp 30
 Db 32573 GCGGCTGATATATCTGCCAGCCGCCGCAACA-----AGTCCGCGGCGAG 32617
 QY 31 AAlaGlnGlyGlnProAlaGlyArgGlnAlaProAlaValAlaGlyValVal 50
 Db 32618 GCTTCAGCAGTGCACCCGCCGCGAATCGCAAAAGTCCGCGATCGTCTTCA 32677
 QY 51 ThrValHisProGln-----ThrValAlaLeuThrValGlnLeuPro 64
 Db 32678 ACAGC-CACACAGAAAGCTGGGGAAGCCATGAACAGCGCAGAGGAGGATTCGC 32736
 QY 65 GlyArgLeuGlnSerLeuArgThrAlaAspValArgAlaGlnValGlyIleIleGln 84
 Db 32737 TTCGCGAGCGAAGACATCCGA-----AACGACGCGAAGCGTTGGAAAC-----CAG 32784
 QY 85 LysArgLeuPheGlnGlnGlySerThrVal-ArgAlaGlyGlnProLeuThrGlnIleAs 104
 Db 32785 CGCAGAAATCTCAAAAACGCGCTGCCGATCGTCAGCC-----AG 32823
 QY 104 pSerSerThrThrArgAlaAsnLeuGlnSerAlaArgAlaGlnLeuAlaThrAlaGlnAl 124
 Db 32824 TTCGCGCGCGTCATCGCATCATCGCTCTCTCAAAAAGATGAGGCGACAGCAAGC 32883
 QY 124 aThrLeuAlaLysAlaAspAlaAspLeuAlaArgThrLysProLeuValAlaAlaGlnAl 144
 Db 32884 GTCAGCAGCGAAGAGCGCAGCAGCATCCACGAAAGCGCAGAGCGAGCGCTGGTAG 32943
 QY 144 aValSerArgGlnGlnLysArgAlaAlaValThrAlaLysArgSerAlaGlnAlaGlyVal 164
 Db 32944 TCCGACG-----GACGACGCTCAGAGCAAAAGTACGGCGGAAATCGCAGC 32988
 QY 164 LlysAlaAlaGlnAlaAlaIleLysSerAla-----GlyIleAsnLeuAs 179
 Db 32989 AACCGCGCTGACAGCAGCGAAGCGGAGAGATTTGCAATCCGCGCGGCGCTTGA 33048
 QY 179 nArgSerArgIleThrAlaProIleSerGlyPheIleGlyGlnSerLysValSerGlnGlu 199
 Db 33049 GGATGCGAGCAGCGAAG-----AAGGGATAGTACAGCTCAGAGTGCAGCTAACAG 33102

QY 199 YThr-----LeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAs 217
 Db 33103 CACTTCGAGTCACTGGCGGAACGCCAAAAGCGTTAAAGGCCGATAGTGCAGCTAA 33162
 QY 217 nProMetCysVal-----As 222
 Db 33163 CGGGAATATACCCGACAGAGATGCAACAGCAGCAAGAAAGGATGTCAGCTTAAGCA 33222
 QY 222 nValThrGlnSerAlaSerGluValMet-----LysLeuArgAr 235
 Db 33223 CGGACCAACAGCAGCATCTGAAATGCTGGCGGCAACGCCAAAGTCGTTAAGGCGACTTA 33282
 QY 235 gGlnIleAlaGlnGlyLysLeuLeuAlaAlaAspGlyValIleAlaValAlaGlyIleLysPh 255
 Db 33283 TGACCTTGCTACCGGAAATATACTGCTCAGAGAGCTATACAGACACA----- 33328
 QY 255 eAspAspGlyThrValLysProGlnLysGlyArgLeuLeuPheAlaAspProValValAs 275
 Db 33329 -----CAAAAAGGAAATGTCAGCTCAGTGTGCAACCAACAG 33366
 QY 275 nGlnSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeu 295
 Db 33367 CGCATCTGAA-----AGGCTT-----GCCGACACCGAAA----- 33397
 QY 295 tProGlyLeuThrValArgValLeuMetAspGlnValAlaValAspAsnAlaPheValVa 315
 Db 33398 -----GCAGTGAAGCAGCTAATGATATATGCG----- 33424
 QY 315 lProGlnAlaValThrArgGlyAlaLysAspThrValMetIleValAsnAlaGlnGlu 335
 Db 33425 -----AATGTCGGGTACTCTTCCGCCGTAAAGGTAAATGTTAAGGC 33465
 QY 335 Y-----GlyMetGluProArgGluVal-----ThrVa 344
 Db 33466 GCTTCATCGGATATTAACCTGACGCGCAAGAAATATTGTACGTTAATCAACACACAT 33525
 QY 344 lAlaGlnGlnGlnGlyThrAsnTrp-----IleValThrSerGlyLeuLysAspGlyLys 362
 Db 33526 GTCAATTACGGGTGCTGCTGTTGTTCAAAATTAGCAACGTACATGCCAGCGGAG 33585
 QY 362 pLysValValValGlnGlyLysSerIleAlaGlyIleThrGlyAlaLysLysValThrPr 382
 Db 33586 TTCTGTGTT-----TCAATTACGTTGATTGTGGCGCGGATTTAACGTGGGTCAAC 33639
 QY 382 cLysGlnTrpAlaSerSerGlu 389
 Db 33640 TCAACAGCAGGATATATCTGAA 33661

Search completed: September 8, 2003, 15:11:57
 Job time : 23898 secs

THIS PAGE BLANK (USPTO)

Result No.	Score	Query Match	Length	DB	ID	Description
1	802.5	39.7	1275	4	US-09-318-352-73	Sequence 73, Apple
2	753	37.3	1251	4	US-09-252-991A-444	Sequence 444, Apple
3	753	37.3	1362	4	US-09-252-991A-417	Sequence 417, Apple
4	701.5	34.7	7866	4	US-09-453-702B-102	Sequence 102, Apple
5	698.5	34.6	1233	4	US-09-252-991A-14019	Sequence 14019, Apple
6	683	33.8	1161	4	US-09-322-911A-14117	Sequence 14117, Apple
7	647	32.0	1344	4	US-09-318-352-3257	Sequence 3257, Apple
8	642	31.8	1245	4	US-09-252-991A-11404	Sequence 11404, Apple
9	619	30.7	1184	4	US-09-252-991A-11488	Sequence 11488, Apple
10	527	26.1	1026	4	US-09-252-991A-11384	Sequence 11384, Apple
11	451.5	22.4	1230	4	US-09-328-352-444	Sequence 444, Apple
12	448	22.2	1350	4	US-09-252-991A-14689	Sequence 14689, Apple

C	45	212.5	10.5	3880	4	US-09-221-017B-1071	Sequence 1071, App
C	44	216.5	10.7	1164	4	US-09-328-352-877	Sequence 877, App
C	43	217	10.7	1464	4	US-09-352-991A-8876	Sequence 8876, App
C	42	220	11.0	1503	4	US-09-328-352-4057	Sequence 4057, App
C	41	221.5	11.0	3150	4	US-09-352-991A-8765	Sequence 8765, App
C	40	221.5	11.0	1353	4	US-09-352-991A-9189	Sequence 9189, App
C	39	237.5	11.8	1245	4	US-09-352-991A-11478	Sequence 11478, App
C	38	239	11.8	465	4	US-09-352-991A-11457	Sequence 11457, App
C	37	244.5	12.2	1197	4	US-09-352-991A-4531	Sequence 4531, App
C	36	245.5	12.2	1170	4	US-09-352-991A-4631	Sequence 4631, App
C	35	245.5	12.6	1011	4	US-09-352-991A-7845	Sequence 7845, App
C	34	253.5	12.6	1101	4	US-09-352-991A-11972	Sequence 11972, App
C	33	256	12.8	1221	4	US-09-352-991A-11937	Sequence 11937, App
C	32	258.5	12.8	2313	4	US-09-352-991A-13042	Sequence 13042, App
C	31	258.5	12.8	1527	4	US-09-352-991A-13006	Sequence 13006, App
C	30	267	13.2	4704	4	US-09-352-991A-4005	Sequence 4005, App
C	29	267	13.2	1398	4	US-09-352-991A-4039	Sequence 4039, App
C	28	267	13.2	1200	4	US-09-352-991A-4039	Sequence 4039, App
C	27	272.5	13.5	1359	4	US-09-328-352-3771	Sequence 3771, App
C	26	277	13.7	1185	4	US-09-352-991A-7642	Sequence 7642, App
C	25	277	13.7	1185	4	US-09-352-991A-7614	Sequence 7614, App
C	24	286	14.2	1830121	4	US-09-643-990A-1	Sequence 1, App
C	23	286	14.2	1830121	4	US-09-557-884-1	Sequence 1, App
C	22	340.5	16.9	1224	4	US-09-352-991A-1946	Sequence 1946, App
C	21	355	17.6	4404	4	US-09-352-991A-1086	Sequence 1086, App
C	20	355	17.6	4404	4	US-09-352-991A-1120	Sequence 1120, App
C	19	370.5	18.4	1008	4	US-09-199-637A-148	Sequence 148, App
C	18	396	19.6	1509	4	US-09-352-991A-7817	Sequence 7817, App
C	17	396.5	19.6	1386	4	US-09-352-991A-7520	Sequence 7520, App
C	16	429	21.2	1347	4	US-09-352-991A-15083	Sequence 15083, App
C	15	429	21.2	1290	4	US-09-352-991A-14670	Sequence 14670, App
C	14	429	21.2	1245	4	US-09-352-991A-14539	Sequence 14539, App
C	13	448	22.2	1368	4	US-09-352-991A-15065	Sequence 15065, App
C	12	449	22.2	1368	4	US-09-352-991A-15039	Sequence 15039, App

ALIGNMENTS

```

RESULT 1
US-09-328-352-73
; Sequence 73, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 73
; LENGTH: 1275
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-73

```

Alignment Scores:

Pred. No.:	4,67e-78	Length:	1275
Score:	802.50	Matches:	181
Percent Similarity:	63.04%	Conservative:	80
Best Local Similarity:	43.72%	Mismatches:	136
Query Match:	39.75%	Indels:	17
DB:		Gaps:	8

Oy 8 A l M e r t g l i a l a l a l e n u a l a l a l a l a l e n u s e r c y s g l y l y s 27
|||:::
Db 52 G C C T T A C T C T G C C C A T T A G C A C A A T A T C C C C T T G T - - - - - C G T T G T A G C A A 106

Oy 28 G L Y E L Y A S P I A A l a c t i n c l y e l g i n p r o a l a g l y a r g e l u a l a p r o a l a p r o v a l y a l 47
|||:::
Db 106 G G C T C C G A T A G A A A C G A C A G C T G C T G C T C A G A A A T C C C G C T T G C A A - - - G T A 166

```

Qy 48 G1yValValThraH1sProGlnThrValAlaLeuThrValGluLeuProGlyArgLeu 67
Db 163 GGTGTATGTTGGTCAACCAAGGTTGACCAAGACCGTTGACGCTTCAAGCCGTAAT 222
Qy 68 GluSerLeuArgThrAlaAspValArgAlaGlnValGlyGlyIleIleGlnLysArgLeu 87
Db 223 TCAGCAATATCAAAATTTCTGAAGTTCTGCTCAACCAAGTGGCGGATTTTAAACGTTTA 282
Qy 88 PheGlnGlnGlySerTyrValArgAlaGlyGlnProLeuTyrGlnIleAspSerSerThr 107
Db 283 TTTCTGTAAGGAAGACTATGTTCTGTAAGGTCAGCGCTTTATGAGCTGACTTGAAGC 342
Qy 108 TyrGlnAlaAsnLeuGlnSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAla 127
Db 343 AACCGTGCAACGTTAGAAATATGCAAAAGCATCATCTCTACCAACCAAGGCAAACTGAGT 402
Qy 128 LysAlaAspAlaAspLeuAlaArgTyrLysProLeuValAlaAlaGlnAlaValSerArg 147
Db 403 TCAGTACGTAACCAAGTTAAATCGTTATTAACCACTGTTCTAGTAATGCTGTCTAAA 462
Qy 148 GlnGlnTyrAspAlaAlaValThrAlaLysArgSerAlaGlyValLysAlaAla 167
Db 463 CAGGAATATGATGACTTATCTGTCAGTCACTGTTGCAAGGCAACAAAGTTCAGAGAGCT 522
Qy 168 GlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsnArgSerArgIleThrAlaProIle 187
Db 523 AAGGCTCAAGTAACAAATGCAAAATGATGATCTGGTATTTCTCAAAATTCGCTCTCAT 582
Qy 188 SerGlyPheIleGlyGlnSerLysValSerGlnGlyThrLeuLeuAsnAlaGlyAspThr 207
Db 583 TCTGGCCAAATCGGTCTGCTTCAAGGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 642
Qy 208 ThrValLeuAlaThrIleArgGlnThrAsnProMetTyrValAsnValThrGlnSerAla 227
Db 643 GACCGGTTGGTAAGGATTCACCAAGTATGATCTTATGATGATGATGATGATGATGAT 702
Qy 228 SerGlnValMetLysLeuArgArgGlnIleAlaGlnGlyLysLeuLeuAlaAlaAspGly 247
Db 703 GCTAAGTATATGCTTACGTCACCAACTAAGTAAAGGCACTTAAATTAACAGTAACAC 762
Qy 248 ValIleAlaValGlyIleLysPheAspAspGlyThrValTyrProGlnLysGlyArgLeu 267
Db 763 ACG---AAAGTAAATTAAGCTTGAAGATGTTCTACCTATCCAAATCGAAGGCGCAAT 819
Qy 268 LeuPheAlaAspProValValAsnGlnSerThrGlyGlnIleThrLeuArgAlaAlaVal 287
Db 820 GCTTCTCTGACGCTTCTGTAACCAAGATACAGGAACATTAACCTTACGTCGCTATTC 879
Qy 288 ProAsnAspGlnAsnIleLeuMetProGlyLeuTyrValArgValLeuMetAspGlnVal 307
Db 880 TCTAACCAGATCAATTTATGCTTCCGGTATGATATACCATCGGCAAAATGTTCCAGGCG 939
Qy 308 AlaValAspAsnAlaPheValValProGlnGlnAlaValThrArg-----GlyValLys 325
Db 940 GTTGTTCCAATGCTTACCTGATTCCTCAACGCTGCAATTAATCTGTTAACCAGACAA 999
Qy 326 AspThrValMetIleValAsnAlaGlnGlyGlyMetGluProArgGluValThrValAla 345
Db 1000 GCTGTAGGATGCTGTTAATGCTAAGGGGTTGAGAGCCGCTCTGTTGAAACCTCT 1059
Qy 346 GlnGlnGlnGlyThrAsnTrpIleValThrSerGlyLeuLysAspGlyAspLeuValVal 365
Db 1060 GGTGTTCAGGACAAACTGATTTGTGACTAAGCGCTTAAAGCGCGGATTAAGTCAATT 1119
Qy 366 ValGlnGlyIleSerIleAlaGlyIleThrGlyAlaLysValThrProLysGlnTrp 385
Db 1120 GTTATGT-----GTTGCCAAAGTTTAAAGAGGCGCAAGATACAGCAAAACCTTAT 1173
Qy 386 AlaSer-----SerGlnAsnGlnAlaAlaIlePro-----GlnSer 397
Db 1174 CAAGCTCAACGACCAACTCTCAAGGTCGACGACCAAAATGCTGCGAAACCGGCTCAATCA 1233
Qy 398 GlyValGlnThrAlaSerGluAlaLysThrAlaSerGluAla 411

```

```

Db 1234 GGT---AAACCTCAAGCAGAAACAGAAAGCACTTCAATCA 1272
RESULT 2
US-09-252-991A-444/C
; Sequence 444, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 444
; LENGTH: 1251
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-444
Alignment Scores:
Pred. No.: 1,15e-72 Length: 1251
Score: 753.00 Matches: 174
Percent Similarity: 58.60% Conservative: 68
Best Local Similarity: 42.13% Mismatches: 125
Query Match: 37.30% Indels: 46
Gaps: 10
US-09-889-756a-2 (1-412) x US-09-252-991A-444 (1-1251)
Qy 8 AlaMetArgAlaAlaAlaLeuAlaAlaValAlaLeu---ValLeuSerSerCysGly 26
Db 1137 GCCATGCGTATGACTGTTCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1078
Qy 27 LysGlyLysAspAlaAlaGlnGlyGlnProAlaGlyArgGlnAlaProAlaProVal 46
Db 1077 AAAAGC-----GAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1054
Qy 47 -----ValGlyValValThraH1sProGlnThrValAlaLeuThrValGlu 62
Db 1053 CAAGCGCGGAGGTCGGGATCGTACCCCTGGAAGCGCAGACCGTGAATACCGAG 994
Qy 63 LeuProGlyArgLeuGlnSerLeuArgThrAlaAspValArgAlaGlnValGlyIle 82
Db 993 CTGCGCGGCGGACCAATGCGTCCGATCGCGAGCGTGGTCCCGCAGTGAAACGGCATC 934
Qy 83 IleGlnLysArgLeuPheGlnGlySerTyrValArgAlaGlyGlnProLeuTyrGln 102
Db 933 ATCTCAAGCGCTGTTCAAGGAAGGACGACCTCAAGCGCGGCGCAGCTTACAG 874
Qy 103 IleAspSerSerThrTyrGlnAlaAsnLeuGlnSerAlaArgAlaGlnLeuAlaThrAla 122
Db 873 ATGACCCCGCCACCTTCAGAGCGGACTACCAAGCGCCCGCAACCTGCTGCAGC 814
Qy 123 GlnAlaThrLeuAlaLysAlaAspAlaAspLeuAlaArgTyrLysProLeuValAlaAla 142
Db 813 CAG-----GAGCAGCGCCAGCGCTACCAAGCTGCTGTCGCGCAC 775
Qy 143 GlnAlaValSerArgGlnGlyTyrAspAlaAlaValThrAlaLysArgSerAlaGluAla 162
Db 774 CAGGCTGTAGCAAGCAGCGATC-----GCCAGCGCCAAATGCC 736
Qy 163 GlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsnArgSerArg 182
Db 735 GCCTACTGACGATCAAGGCGGCGGTGAGACGCGCGGATCAACCTGCGCTACACCAAG 676
Qy 183 IleThrAlaProIleSerGlyPheIleGlyGlnSerLysValSerGlnGlyThrLeuLeu 202

```

```

Db      675 GTGCTGTCGGCATCTCCGGCCGATCGCGCTTCGCGGTGACCGAAGCGCCCTGCTG 616
Qy      203 Asn1agiYAspThrThrValIleuAlaThrIleArgInThrAsnProMetTyrValAsn 222
Db      615 ACCAAGCGGCGAGCGCAAGCGATGCCGCGCAACAGCTCGACCCGATCTACGTCGAC 556
Qy      223 ValThrGlnSerAlaSerGluValMetIleuArgArgGlnIleIleGluGlyLysIleu 242
Db      555 GTCAACCGACCGCTCCACCGCGCTGCTGCGCGCGCAACTGCGACGCGCGCAAGTTG 496
Qy      243 ---LuuAlaIleAspGlyValIleAlaValGlyIleLysPheAspAspGlyThrValTyr 261
Db      495 GAGCGCGCGCGCGCAACGCGCGCAAGGTCCTCCGAGAGTGGAGAGCGTACCAATAC 436
Qy      262 ProGluIysGlyArgLeuPheAlaAspProValValAsnGlnSerThrGlyGlnIle 281
Db      435 CCGCTGAGAGGTCCGCTCGAATTCCTCCGAGTTTCGTCGACGAAAGGACCGGCTCGCTC 376
Qy      282 ThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetProGlyLeuTyrValArg 301
Db      375 ACCATCCGCGCGCTGCTCCCAACCGCAACAGAGCTGTCGCGGCGATGTCGTTAC 316
Qy      302 ValLeuMetAspGlnValAlaValAspAsnAlaPheValValProGlnGlnAlaValThr 321
Db      315 GCGCAGTTGCAAGAGCGCTCAACGCAAGGCAATCTCTGCTCGCGCAAGGCGGTGACC 256
Qy      322 ArgGlyAlaLysAsp-----ThrValMetIleValAsnIleGlnGlyMetGluPro 339
Db      255 CGCAGCTCAAGAGCGCGCAAGCTACCGCGCTGCTGTCGTCGTCGTCGTCGTCGTCGTCG 196
Qy      340 ArgGluValThrValAlaGlnGlnGlnGlyThrAsnTyrIleValThrSerGlyLeuLys 359
Db      195 CGGGATGTCAGAGCGCGCAAGCGGTGATCGCGCAAAATGCTGTTACCGAAGGCTGTAAC 136
Qy      360 AspGlyAspLysValValValGluGlyIleSerIleAla-----GlyIleThrGlyAla 377
Db      135 GCGCGCGCAAGATCATTAACGAAAGCGCTGCGATTCGTCGACCGCGGTGTCGAG---GTG 79
Qy      378 LysLysValThrProLysGluTyrPalaSerSerGlnAsnGlnAlaIleAlaProGlnSer 397
Db      78 AAGACCTGCGCGCGCAAGATGTCGCTCGCGCAAGAGCGCGCGCTCGC----- 25
Qy      398 GlyValGlnThrAlaSerGluAlaLysThrAlaSerGlu 410
Db      24 -----GCGAAAACGACGACGACG 7

RESULT 3
US-09-252-991A-417
; Sequence 417, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252, 991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 31142
; SEQ ID NO 417
; LENGTH: 1362
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-417

Alignment Scores:
Pred. No.: 1,32e-72 Length: 1362
Score: 753.00 Matches: 174
Percent Similarity: 58.60% Conservative: 68
Best Local Similarity: 42.13% Mismatches: 125

```

```

Query Match: 37.30% Indels: 46
DB: 4 Gaps: 10
US-09-889-756a-2 (1-412) x US-09-252-991A-417 (1-1362)

Qy      8 AlaMetArgAlaAlaAlaLeuAlaAlaValAlaLeu---ValLeuSerSerCysGly 26
Db      226 GCCATGCGTGTACTGTTCCGCGCTGCTGTCGATTCGCGCTTTCGCGGTGCGGA 285
Qy      27 LysGlyLysAspAlaAlaGlnGlnGlyGlnProAlaGlyArgGluAlaProAlaProVal 46
Db      286 AAAAGC-----GAGGCGCGCGCGCGCG 309
Qy      47 -----ValGlyValAlaThrValHisProGlnThrValAlaLeuThrValGlu 62
Db      310 CAAGCGCGGAGGTCCGCGATCTGATCCCTGGAAGCGCAAGCGGATCTGAAATACGAG 369
Qy      63 LeuProGlyArgLeuGlnSerLeuArgThrAlaAspValArgAlaGlnValGlyIle 82
Db      370 CTGCGCGCGCGCGCAACATGCGTTCCGATCGCGAGGTGCGTCCCGAGTGAACGCGATC 429
Qy      83 IleGlnLysArgLeuPheGlnGlnGlnGlySerTyrValArgAlaGlyGlnProLeuTyrGln 102
Db      430 ATCTCAAGCGCTGTCAGAGAGGCGAGCGTCAAGCGCGGCGAGAGCTTACAG 489
Qy      103 IleAspSerSerThrTyrGluAlaAsnLeuGlnSerAlaArgAlaGlnIleValAlaThrAla 122
Db      490 ATGACCGCGCGCGCACTACGAGCGCGATACCAAGCGCGCAACCTGCTTCGAC 549
Qy      123 GlnAlaThrLeuAlaLysAlaAspAlaAspLeuAlaArgTyrIleProLeuValAlaAla 142
Db      550 CAG-----GAGCAGCGCGCGCGTACCAAGCTGCTGTCGCGAC 588
Qy      143 GlnAlaValSerArgGlnGlnTyrAspAlaAlaValThrAlaLysAspSerAlaGluAla 162
Db      589 CAGCGCGTGAAGCAGCAGCTAC-----GCCGACGCCAATGCC 627
Qy      163 GlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsnArgSerArg 182
Db      628 GCTTACTCGTAGTTCAGAGCGCGGTGAGCAGCGCGGATCACTCGCTACACCAAG 687
Qy      183 IleThrAlaProIleSerGlyPheIleGlyGlnSerLysValSerGlnGlyThrLeuLeu 202
Db      688 GTGCTGCGCGCATCTCCGCGCGCATCGCGCTTCGCGGTGACCGAAGCGCGCTGCTG 747
Qy      203 AsnAlaGlyAspThrThrValIleuAlaThrIleArgGlnThrAsnProMetTyrValAsn 222
Db      748 ACCAAGCGGCGAGCGCAACGCGATGGCGCACCGTCAACAGCTCGACCCGATCTACGTCGAC 807
Qy      223 ValThrGlnSerAlaSerGluValMetLysLeuArgGlnIleAlaGlnGlyLysLeu 242
Db      808 GTACACCGACCGCTCCACCGCGCTGCGCTGCGCTGCGCGCACTGCGACGCGCGCACTG 867
Qy      243 ---LuuAlaIleAspGlyValIleAlaValGlyIleLysPheAspAspGlyThrValTyr 261
Db      868 GAGCGCGCGCGCGCAACGCGCGGAGGAGTCTCCCTGAAGGTGAGAGCGGTAGCAATAC 927
Qy      262 ProGluIysGlyArgLeuPheAlaAspProValValAsnGlnSerThrGlyGlnIle 281
Db      928 CCGCTGAGAGGTGCTCGAATTCCTCCGAGTTTCGTCGACGAAAGCGCGCTGCTGCTC 987
Qy      282 ThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetProGlyLeuTyrValArg 301
Db      988 ACCATCCGCGCGCTGCTCCCAACCGCAACAGAGCTGCTGCGCGCATGTCGTTAC 1047
Qy      302 ValLeuMetAspGlnValAlaValAspAsnAlaPheValValProGlnGlnAlaValThr 321
Db      1048 GCGCAGTTGAGAGAGCGCTCAAGCAGAGGCAATCTGCTCGCGCAAGCGGTGAC 1107
Qy      322 ArgGlyAlaLysAsp-----ThrValMetIleValAsnIleGlnGlyMetGluPro 339
Db      1108 CGGACCTCAAGGCGCGAGCTACCGCGCTGCTGTCGTCGTCGTCGTCGTCGTCGTCG 1167

```

[illegible]

QY	7	LysalMetAcTAgAlaAlaAlaLeuAlaAlaAlaValAlaLeuValLeuSerSerCysGly	26
Db	6823	AAATTAAGAAAGTATATAGCAACATCTGTAGTGGCAATGCTGCTTATGCTGTGT--	6767
QY	27	LysGlyGlyAspAlaAlaGlnGlyGlyGlnProAlaGlyArgGlyAlaProAlaProVal	46
Db	6766	-----GATTAATAGCCAAAGTAAACATTATCCCGTCAGAAACAGAA-----	6725
QY	47	ValGlyValValThrValHisProGlnThrValAlaLeuThrValGluLeuProGlyArg	66
Db	6724	GTCGGCGGTGTTCACGTCAATCTCAACCCGGTCTCGTAGTCAGTGAATTAAACCGACGC	6665
QY	67	LeuGluSerLeuAqThrAlaAspValArgAlaGlnAlaGlyGlyIleIleGlnLysArg	86
Db	6664	ACCAAGTGTGGCTCGACGTCCGAAGTACGTCCGAGGTGGGGAAATTATCCAAACACGC	6605
QY	87	LeuPheGlnGlnGlySerTyValArgAlaGlyGlnProLeuTyArgIleAspSerSer	106
Db	6604	TTATTTAAGGAAGTGATCTGTCTCAAGCTGGACGCCCTCTACAGATTGATGCGCGC	6545
QY	107	ThrTyrgLualaAsnLeuGlnSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeu	126
Db	6544	AGTTATACGCTGCATGAGTAAAGTACGTCCGGACACATTACAACAACACAGCGA---CTG	6488
QY	127	AlaLysAlaAspAlaAspLeuAla--ArgTyLysProLeuValAlaAlaGluAlaVal	145
Db	6487	GTAAGAAGCCGATTCGCCAAGAACCCACGGTATTCGCCAGCTGGGAAGAACGCGTGT	6428
QY	146	SerArgGlnGlnTyArgAspAlaAlaValThrAlaLysArgSerAlaGluAlaGlyValLys	165
Db	6427	TCACACACGAGTGTGATGATGCTCACTCACTCTGTCACACAGTAAAGCCAGTGTAGCG	6368
QY	166	AlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsnArgSerArgIleThrAla	185
Db	6367	GCAGAAAAAGCCGACACGAAACGCGCGCATTTATCTTGACCTGCAGCACCGTAAACCGCA	6308
QY	186	ProIleSerGlyPheIleGlyGlnSerLysValSerGlnGlyThrLeuLeuAsnAlaGly	205
Db	6307	CCGATTTCCAGGCGCCATTTGGCATTTCTCGGTAAACCCCTGGCGACCTGCACCGCGCTCG	6248
QY	206	AspThrThrValLeuAlaThrIleArgGlnThrAsnProMetTyValAsnValThrGln	225
Db	6247	CAGATACAGCGTTAACACTATTCTGCTGTGATTCAAATGATATGTCAGACTCACTCGC	6188
QY	226	SerAlaSerGluValMetLysLeuArgArgGlnIleAlaGlnGlyLysLeuLeuAlaAla	245
Db	6187	TCCAGTGTGCATTTATTATTCAGTCTCCGTAAACG-----TCACGTGGCAACAC	6140
QY	246	AspGlyValIleAlaValGlyIleLysPheAspAspGlyThrValTyProGlnLysGly	265
Db	6139	AGTGAACACCAAGCAGTCTCACTATTCTCGAAGATGGCACACCTTCACAGGAAAAAGGG	6080
QY	266	ArgLeuLeuPheAlaAspProValValAsnGlnSerThrGlyGlnIleThrLeuArgAla	285
Db	6079	CGCTCGAAGCTCACCGAAGTCGCTGTAGATGAGTCTACCGGTCTGGTACATTACGGGCGC	6020
QY	286	AlaValProAsnAspGlnAsnIleLeuMetProGlyLeuTyValArgValLeuMetAsp	305
Db	6019	ATTTTCCCAATCCACACAGCACTATTATTCACGGGAATGTTTGTCCGCTCGCTGCAT	5960
QY	306	GlnValAlaValaAspAsnAlaPheValValProGlnGlnAlaValThrArgGlyAlaLys	325
Db	5959	GAAAGCGGTGAAGAAAGCGTATTCTCGCGCCGCAACAGGGCGTACACCGCGGATCTTAA	5900
QY	326	Asp-----ThrValMetIleValAsnAlaGlnGlyMetGluProArgGluValThr	343
Db	5899	GGCATATGCACCTGCGCTGTGTGTAATTAAGACATTAAGATGAGCGAACCCTCGAA	5840
QY	344	ValAlaGlnGlnGlnGlyThrAsnThrIleValThrSerGlyLeuLysAspGlyAspLys	363
Db	5839	ACGGAGAAACATATGGTGAATAATGGCTGTGCTTAACGGCTCTGCACACGGCGGACCGA	5780

```

Qy 364 ValValValGluGlyLeuSerIleAlaGlyIleThrGlyAlaValValValThrProLys 383
Db 5779 CTGATGTGTAAGT-----TCTGCCAAGTCACTTCAGGACGAGCTCAAGCTGTT 5726
Qy 384 Glu 384
Db 5725 GAA 5723

RESULT 5
US-09-252-991A-14019
; Sequence 14019, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 14019
; LENGTH: 1233
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14019

Alignment Scores:
Pred. No.: 1e-66 Length: 1233
Score: 698.50 Matches: 164
Percent Similarity: 62.66% Conservative: 81
Best Local Similarity: 41.94% Mismatches: 130
Query Match: 34.60% Gaps: 17
Db: 4 Gaps: 8

US-09-889-756a-2 (1-412) x US-09-252-991A-14019 (1-1233)
Qy 10 ArgAlaAlaAlaLeuAlaAlaAlaValAlaLeuValLeuSerSerCysGlyValGlyGly 29
Db 94 AGGATCGGGCGGTGGCTATGGCCATCGC-----TTGGCGGGTGTGGGCC-GGGCGA 146
Qy 30 AspAlaIaGlnGlyGlnProAlaGlyArg-GluAlaProAlaProAlaValGlyVa 49
Db 147 AGAGCGCA-----GGAGCGCGCGAATGATGTTGCCG--GTGAGGT 188
Qy 49 ValThrValHisProGlnThrValAlaLeuThrValGluLeuProGlyArgLeuGln 69
Db 189 CTGACGCGTCCAGGCGCCGCTGGCGGTGAGTTCGGAATCGCTGGCGGATCGAAC 248
Qy 69 rLeuAlaGlyThrAlaAspValAlaGlnValGlyGlyIleIleGlnLysArgLeuPheG 89
Db 249 GGTGCGGTGCGCGAGGTGCGCGCGGTGCGCGCGCATGCTGCGCGAAGCGCTTCA 308
Qy 89 ngLysSerTyValArgAlaGlyGlnProLeuTyArgGlnIleAspSerSerThrTyrg 109
Db 309 GGAAGGCGCGCAGCTCAAGCTGCGCATCTGCTGTTCATGATCGGACCGCTGAA 368
Qy 109 uAlaAsnLeuGlnSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaLysAl 129
Db 369 GGTGCGGTGCTCCGCGCGCGAGGTGAGCTGCGCGCGAACCAGCGGTGCTGTTCAGGC 428
Qy 129 aAspAlaAspLeuAlaArgTyArgProLeuValAlaAlaGlnAlaValSerArgGln 149
Db 429 GCAAGCGCGGTGCTGCTGCTACAGCGCGGTGAGATCGACGCGCTCAGCGACGAGA 488
Qy 149 uTyAspAlaAlaValThrAlaLysArgSerAlaGlnAlaGlyValLysAlaAlaGlnAl 169
Db 489 CTTCGATACCGCCACCGCCAGCTGCGCGAGCGCGCGGCGCAACCGCTGCGCGCAAGG 548
Qy 169 aAlaIleLysSerAlaGlyIleAsnLeuAsnArgSerArgIleThrAlaProIleSerG 189

```

```

Db 549 CGACCTGGAACCGCGCGCTGGAACCTCGGCTACGCTCGGATCTCGCGATCTCCGG 608
Qy 189 yPheIleGlyGlnSerTyValSerGlyThrLeuLeuAsnAlaGlyAspThrThrVa 209
Db 609 GCGCATCGCGCGCGCTGCTGACCGAGGCGCGCTGTGTGGCGAGCGCGAGCGGAGCT 668
Qy 129 uAlaAlaThrIleArgGlnThrAsnProMetTyValAsnValThrGlnSerAlaSerG 229
Db 669 GATGCGCGCATCAGAGATTCATCGATCTATGCGGATTCACCCAGACCGCGCGCA 728
Qy 229 uValMetLysLeuAlaArgGlnIleAlaGlyLysLeuLeuAlaAlaAspGlyVal 249
Db 729 GCGCTCGCGCTCGCGAGCGCGCTGGAAGAACCTTGGCGCGCGCGAGCGAGCCAG 786
Qy 249 eAlaValGlyIleLysPheAspArgGlyThrValTyProGlnLysGlyArgLeuPhe 269
Db 787 -GGCTGACCTTCGGGTC--GAGGAGCGCTTCAGACCGCGCGCGCTGTCAGTT 842
Qy 269 eAlaAspProValValAsnGlnSerThrGlyGlnIleThrLeuArgAlaAlaValProAs 289
Db 843 CGCGACGTGCGCGGTGATCGCGTACCGGCGAGATCCCTCGCGCGCAAGTCCGCA 902
Qy 289 nAspGlnAsnIleLeuMetProGlyLeuTyValArgValLeuMetAspGlnValAlaVa 309
Db 903 CCGCGAGCGGCTCTGCTGCGCGCATGACGTGCGCGTACGTAACGCCCGAGGCGATCGA 962
Qy 309 IaAspAlaAlaPheValValProGlnIleAlaValThrArgGlyAlaLysAspThr----- 327
Db 963 CAACGAGCGATCTCTGCTGCGCGCAAGCGCGCGTGCACCGCTCCAGCGAGCGCGCCA 1022
Qy 328 -ValMetIleValAlaGlnGlyIleGlyMetGluProArgGluValThrValAlaGln 347
Db 1023 GGTGATGTGTGTGGCGCGCGAGCGCGCGCGAGTCCGCGAGCTGTCGATCCGCGTCA 1082
Qy 347 ngGlnTyThrAsnTrpIleValThrSerGlyLeuLysAspGlyAspLysValValG 367
Db 1083 GCAAGGTGCGCGTGCAGATCACGAGGCGCTGAGCGCGGTGACCGGCTATGTCGG 1142
Qy 367 uGlyIleSerIleAlaGlyIleThrGlyAlaLysValThrProLys---GluTrpAl 386
Db 1143 CGCG-----CTGCTGCGGTGCGAGCGCGGCGGTGAAGATCTGTCCGAAGCGGATGTC 1196
Qy 386 aSerSerGluAsnGlnAlaAlaAlaProGln 396
Db 1197 CAGGCGCAAGCCCACTGCTGCGCGCA 1227

RESULT 6
US-09-252-991A-14117/c
; Sequence 14117, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 14117
; LENGTH: 1161
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14117

Alignment Scores:
Pred. No.: 4.5e-65 Length: 1161
Score: 683.00 Matches: 159
Percent Similarity: 62.60% Conservative: 77

```



```

Db      1174 AAGCGGCGTGTCCGCGACGCGCGACGCGCGACGCCCTCTACCGCTGTGGCGCTGCGCG 1233
QY      411 Ala 411
        |||
Db      1234 GCG 1236

RESULT 9
US-09-252-991A-11488/c
; Sequence 11488, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107136.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11488
; LENGTH: 1104
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11488

Alignment Scores:
Pred. No.:      4,04e-58      Length:      1104
Score:          619.00      Matches:      145
Percent Similarity: 57.64%      Conservative: 70
Best Local Similarity: 38.87%      Mismatches: 148
Query Match:    30.66%      Indels:      10
DB:             4           Gaps:      4

US-09-889-756A-2 (1-412) x US-09-252-991A-11488 (1-1104)
QY      20 LeuValLeuSerSerCysGlyLySgLyGlyValAspAlaIaGlnGlyGlnProIaGly 39
Db      1104 CTATTCTCGTGGGCGGTGGAGAGACAGCGGAC-----GCCGGG 1066

QY      40 ArgGlyIaPProIaIaProVal---ValGlyValValThrValHisProGlnIthValAla 58
        ::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1065 AAGACTGGGAGGCGCCCGCGCGAGGTGGCGCTGATCTCGGCAGCGCGCGCGCTATGGC 1006

QY      59 LeuThrValGluLeuProGlyArgGlnGlnSerLeuArgThrAlaAspValArgAlaGln 78
        ::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1005 ATCCACCGACGAGCTGCCCCGAGCGCTGTGAMCGGTACCGCAGGCTGAAATGCGGCGCGC 946

QY      79 ValGlyGlyIleIleGlnIlySargLeuPheGlnGlnGlySerTyValArgAlaGlyGln 98
        ::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      945 GTCCGCGGCATCGTCACCCCGTGGCTGTGACGAGGAAGCCAGAGCGTCGCCGCGCGCAC 886

QY      99 ProLeuTyGlnIleAspSerSerThrTyGlnAlaAsnLeuGlnSerAlaArgAlaGln 118
        ::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      885 GTGCTGTCTCCAGATCGACCCCTGCGCCCTTGAAAGCGCGCCCTTGACATCAGCGCGCGCC 826

QY      119 LeuAlaThrAlaGlnAlaThrLeuAlaIlyAlaAspAlaAspLeuAlaArgTyIlySPro 138
        :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      825 CTGGCCCGCGCGCGCGCGACCGACGCGCGCGCGCGCGCGACCAAGCTCAAGCGCTACGCCGAC 766

QY      139 LeuValAlaIaIaGlnAlaValaIserArgGlnGlnIuTyAspAlaIaValaThrAlaIlySArg 158
        :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      765 CTGATCAAGACCGCGCCCATCGAGCAAGCGAGTACACCGAAGGCGAGACCGACCGCGC 706

QY      159 SerAlaGluIaGlyValIlySAlaAlaGlnAlaIaIleuSeralIaGlyIleAsnIleu 178
        :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      705 CAGGCGCCGCGCGAGATGCGCTCGGCCAAGCGCGACGTGAGAGCGCGCGCGCTGCGCGCTG 646

QY      179 AsnArgSerArgIleThrAlaProIleSerGlyPheIleGlyGlnSerIlySValSerGln 198
        :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      645 GGCTAAGCGCACGATCACCGCGCGCATGACGCGCGCGCGCGCGCGCGCGCTGTGTCACCGAA 586

```

US-09-889-756A-2 (1-412) x US-09-252-991A-11384 (1-1026)

```
QY 1199 |||ThyThLeuLeuAenAlaAGlyAspThrThrValLeuAlaThrIleAArgIlntrAAspPro 218
Db 585 GGGCGCGTGGTGGCGGAGACTCGCCGACACCTTGTGACCCCGCTCGAGCAAGATCGATCCG 526
QY 219 MetTyValAsnValIThrGlnSerAlaSerGlnValMetLysLeuAArgArgGlnIleAla 238
Db 525 ATTCATCGGAACCTTCTCCACAGCCGGCGGGAAGTCCGCCCATGACGGGGCGATCCGC 466
QY 239 GlnGlyLysLeuLeu---AlaAlaAspGlyValIleAlaValGlyIleLysSphAspAsp 257
Db 465 GAAGGCCAGGGAAGGCTGTCGCGACAGAGACATCCCGCTGGCCCTGTGTCGCGCCGAC 406
QY 258 GlnTyThValTyProGlnTyGlyArgLeuLeuPheAlaAspProValValAsnGlnSer 277
Db 405 GGCAGCGCATGCCGCTGCGCGCGGAGACTGTCTTTCGACCTCGGGGCTGCACCCCGGC 346
QY 278 ThrGlyGlnIleThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetProGly 297
Db 345 ACCGACACCATCGGCATCGGTCCCTGTTCCGATCCGATCGCATGCGCATGCCCGGC 286
QY 298 LeuTyValArgValLeuMetAspGlnValAlaValAsnAsnAlaPheValValProGln 317
Db 285 GGCCTACGCGAGGTGCGCGCTGCACGCGCGGTGAACCCGAGGGATCACCGTCCGCGC 226
QY 318 GlnAlaValThrArgGlyAlaLysAspThrValMet---IleValAsnAlaGlnGly 336
Db 225 GACGCGCGTATCGCTTACCGCCAGTCCGCGGTGTCAAGTGTCAACCCACAGGGCTTG 166
QY 337 MetGluProArgGluValThrValAlaGlnGlnGlnGlyThrAsnTrpIleValThrSer 356
Db 165 GTGGAAGACGTGGAGGTCCGCGCGACACCCCTGACAGGGCGCGCATCGATCATCAGCAGC 106
QY 357 GlnTyLeuLysAspGlyAspLysValValValGlnGlyLysSerIleAlaGlyIleThrGly 376
Db 105 GGGCTCAAGGGCGCGCGAGCGGTGATGTGTGAGAAGCCGCCACGATGCCCGCGCTCC 46
QY 377 AlaLysLysValIThrProLysGlnTrpAlaSerSerGlu 389
Db 45 AGCGTCCAGGCGGTGTCTCCGACGCCGCGCACGCGCCGAC 7

RESULT 10
US-09-252-991A-11384
; Sequence 11384, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11384
; LENGTH: 1026
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11384

Alignment Scores:
Pred. No.: 4e-48 Length: 1026
Score: 527.00 Matches: 129
Percent Similarity: 55.16% Conservative: 58
Best Local Similarity: 38.05% Mismatches: 138
Query Match: 26.10% Indels: 14
DB: 4 Gaps: 3
```



```

QY 75 ValArgAlaGlnValGlyIleIleGlnLysArgLeuPheGlnGlySerThrVal 94
Db 2 GTGCGGGGGCGCGTCCGCGCATCGTCCAGCCGCTGTCAGGAAGGCGCAGAGCTC 61
QY 95 ArgAlaGlyGlnProLeuThrGlnIleAspSerThrThrGluAlaAsnLeuGlnSer 114
Db 62 CGCCCGCGCGACCGTGTCTGTCAGATCGACCTCGCCCTTGAAAGCGCGCTCGACATC 121
QY 115 AlaArgAlaGlnLeuAlaThrAlaGlnAlaLeuAlaLysAlaAspAlaAspLeuAla 134
Db 122 AGCCCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 181
QY 135 ArgThrLysProLeuValAlaAlaGlnAlaValSerArgGlnLysThrAspAlaVal 154
Db 182 CGCTACCGCGCGCGCATCGATCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 241
QY 155 ThrAlaLysArgSerAlaGlnAlaGlyValLysAlaAlaGlnAlaAlaIleLysSerAla 174
Db 242 ACCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 301
QY 175 GlyLeuAsnLeuAsnArgSerArgIleThrAlaProIleSerGlyPheIleGlyGlnSer 194
Db 302 CGCTCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 361
QY 195 LysValSerGlnGlyThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIleArg 214
Db 362 CTGCTACCGGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 421
QY 215 GlnThrAsnProMetThrValAsnValThrGlnSerAlaSerGluValMetLysLeuArg 234
Db 422 CAGATCATCGATCGATCGATCGATCGATCGATCGATCGATCGATCGATCGATCG 481
QY 235 ArgGlnIleAlaGlnGlyLysLeuLeu--AlaAlaAspGlyValIleAlaGlyIle 253
Db 482 CGGCGCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 541
QY 254 LysPheAspAspGlyThrValThrProGlnLysArgLeuLeuPheAlaAspProVal 273
Db 542 GTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 601
QY 274 ValAsnGlnSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAspGlnAsnIle 293
Db 602 GTGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 661
QY 294 LeuMetProGlnLeuThrValArgValLeuMetAspGlnValAlaValAspAlaPhe 313
Db 662 TTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 721
QY 314 ValValProGlnGlnAlaValThrArgValAlaLysAspThrValMet--IleValAsn 332
Db 722 ACCGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 781
QY 333 AlaGlnGlyMetGluProArgGlnValThrValAlaGlnGlnGlnGlyThrAsnTrp 352
Db 782 CCACAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 841
QY 353 IleValThrSerGlyLeuLysAspGlyAspLysValValAlaGlnGlyIleSerIleAla 372
Db 842 ATATCATCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 895
QY 373 GlyIleThrGlyAlaLysValThrProLysGlnLysThrAlaSerSerGluAlaGlnAla 392
Db 896 -----CAGATGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 925
QY 393 AlaAlaProGlnSerGlyValGlnThrAlaSerGlnAlaLysThrAlaSerGluAla 411
Db 926 GTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 982

```

RESULT 11
 US-09-328-352-444
 ; Sequence 444, Application US/09328352
 ; Patent No. 6562958
 ; GENERAL INFORMATION:

```

; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAIUMANNIT FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 444
; LENGTH: 1230
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
; US-09-328-352-444

Alignment Scores:
Pred. No.: 9,32e-40 Length: 1230
Score: 451.50 Matches: 131
Percent Similarity: 51.45% Conservative: 82
Best Local Similarity: 31.64% Mismatches: 180
Query Match: 22.36% Indels: 21
DB: Gaps: 7

US-09-889-756a-2 (1-412) x US-09-328-352-444 (1-1230)
QY 4 TyrAlaPheLysAlaMetArgAlaAlaAlaLeuAlaAlaValAlaLeuValLeuSer 23
Db 31 TTTCACATGCTGCTCCATCTTGTGCTATTTAGCAACCGGTGGACGATTATGTTG--- 87
QY 24 SerCysGlyLysGlyValAspAlaAlaGlnGlyGlnProAlaGlyArgGluAlaPro 43
Db 88 ---TTACATGAAATGCGCGATGCGAAAGCTGCACCAACCGCTCCCAACAGCTGCTACT 144
QY 44 AlaProValAlaGlyValValThrValIleProGlnThrValAlaLeuThrValGluLeu 63
Db 145 GTTGATGTACCCCGAGTAGTACG-----AAACCATTCACGATTGGCAAGATAT 195
QY 64 ProGlyArgLeuGlnSerLeuArgThrAlaAspValArgAlaGlnValGlyIleIle 83
Db 196 TCGGTCGCTTTAAGCAATGATCAAGTTGATATTCGGCTCAGATTGAGGAAACTT 255
QY 84 GlnLysArgLeuPheGlnGlnGlySerThrValArgAlaGlyGlnProLeuThrGlnIle 103
Db 256 ATTCGCGTACATTTCAGAGTAGAGGACCTCGTTAAAAAGGTGATTTACTTTTCACATC 315
QY 104 AspSerSerThrThrGlnAlaAsnLeuGlnSerAlaArgAlaGlnLeuAlaThrAlaGln 123
Db 316 GACCTCGCTCTTTAAGCAATGATCAAGTTGATATTCGGCTCAGATTGAGGAAACTT 375
QY 124 AlaThrLeuAlaLysAlaAspAlaAspLeuAlaArgThrLysProLeuValAlaAlaGlu 143
Db 376 GCACAGGTAAATATACCGCAAGCATCTTGGCGTATTCACGTCATTCAGAGTAAT 435
QY 144 AlaValSerArgGlnGlnThrAspAlaAlaValThrAlaLysArgSerAlaGluAlaGly 163
Db 436 GCTGTTTCGCGCAACATGATTTAGCCGAATATGATGACGTTTCAGGAAATGCTTAC 495
QY 164 ValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsnArgSerArgIle 183
Db 496 CTACAGCGCGCTAGACCTGCTGCTCCATCTGCGTTTAATTTAGAAATACACCGTATT 555
QY 184 ThrAlaProIleSerGlyPheIleGlyGlnSerLysValSerGlnLysThrLeuLeuAsn 203
Db 556 ACAGACATGCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 615
QY 204 AlaGlyAsp---ThrThrValLeuAlaThrIleArgGlnThrAsnProMetThrValAsn 222
Db 616 GCAGGTACAGCGCGCAAGGTTTAAACAGTTAGTGTCTGTATCAGCCATATGAGATCT 675
QY 223 ValThrGlnSerAlaSerGluValMetLysLeu-----ArgArgGlnIleAlaGlu 239
Db 676 TTTCATGTTGATGAAACAACTTACCTGAATATATCATGATATACGCTATTCAGACAA 735
QY 240 GlyLysLeuLeuAlaAlaAspGlyValIleAlaValGlyIleLysPheAspAspGlyThr 259

```


QY 406 LysThrAlaSerGluAlaGlu 412
Db 165 AAGGTGGCGGCTCCCAAGAC 145

RESULT 13

US-09-252-991A-15065
Sequence 15065, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

PRIOR FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 15065

LENGTH: 1368

TYPE: DNA

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-15065

Alignment Scores:

Pred. No.: 2,66e-39 Length: 1368

Score: 448.00 Matches: 133

Percent Similarity: 51.11% Conservative: 75

Best Local Similarity: 32.68% Mismatches: 177

Query Match: 22.19% Indels: 22

DB: 4 Gaps: 9

US-09-889-756a-2 (1-412) x US-09-252-991A-15065 (1-1368)

QY 14 LeuAlaAlaAlaValAlaLeuValLeuSerSerCysGlyLysGlyAlaAlaAlaGln 33

Db 172 CTCGCACTCTGCGCGCGGTCCTGCTGAGCGCTGCGCAAGGCCCGCAACACCCCA 211

QY 34 GlyGlyGlnProAlaGlyArgGluAlaProAlaProValValGlyValAlaThrValHis 53

Db 232 GGC-----ATGCGCGCGCCCAAGGTCAAGCTCGCGCGCAAGTCATC 270

QY 54 ProGlnThrValAlaLeuThrValGlnLeuProGlnValArgLeuSerLeuArgThrAla 73

Db 271 GAACAACTGCTGAACAGAGTGAAGATTCAACCGCGCTGAGAGGCCCGAGATCGGTG 330

QY 74 AspValArgAlaGlnValGlyGlyLeuLeuGlnLysArgLeuPheGlnGlnGlySerTyr 93

Db 331 GAGCTGGCGCGCGGCTGCTGCTGATGACCGCGCGCTTCCATGAAAGCGGACATG 390

QY 94 ValArgAlaGlyGlnProLeuTyrGlnIleAspSerSerThrTyrGluAlaAsnLeuGln 113

Db 391 GTGAAGAAAGCACTGCTGCTTCCAGATGACCGCGCGCTTCCAGAGCGGAGTCAAG 450

QY 114 SerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaLysAlaAspAlaAspLeu 133

Db 451 CGCTGCAACCCAGGAGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 510

QY 134 AlaArgTyrLysProLeuValAlaAlaGluAlaValSerArgGlnGlnTyrAspAlaAla 153

Db 511 CACGGCGCGCAAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 570

QY 154 ValThrAlaLysArgSerAlaGluAlaGlyValLysAlaAlaGlnAlaAlaIleLysSer 173

Db 571 ACCACCG 630

QY 174 AlaGlyIleAsnLeuAsnArgSerArgIleThrAlaProIleProIleGlyPheIleGlyGln 193

Db 631 GCGCGCGCTGAACCTTGAACCTTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 690

QY 194 SerLysValSerGlnGlyThrLeuAsnAlaGlyAspThrThrValLeuAlaThrIle 213

Db 691 GCCAGATCACCCCGCAACCTGCTCAACTCCGGGAA--ACCTGCTCACACCTCTG 747

QY 214 ArgGlnThrAspProMetTyrValAsnValThrGlnSerAlaSerGluValMetLys--- 232

Db 748 GTACGACACGACAGGTCTACGCTTACTTCGACCGCGCAAGACCGCTGCTCCCAAGTAC 807

QY 233 -----LeuArgArgGlnIleAlaGlnGlyLysLeuAlaAlaAspGlyValIleAla 250

Db 808 GTGACCTGGCCCGCAG-----GCCGTCGGGACACCGCGCGCGCGCGCGCGCTAC 861

QY 251 ValGlyIleLeuPheAspArgGlyThrValTyrProGlnLysGlyArgLeuPheAla 270

Db 862 CTCGCGCTGACGACGACGACGAC-----CCGACCTGGCGCGCTGACCTTCTC 915

QY 271 AspProValValAsnGlnSerThrGlnIleThrLeuArgAlaAlaAlaProAsnAsp 290

Db 916 GACACACGATCAACCCCGGCTCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 975

QY 291 GlnAsnIleLeuMetProGlyLeuTyrValArgValLeuMetAspGlnValAlaValAsp 310

Db 976 AAGGGGAGTTACCCCGGCGCTCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1035

QY 311 AsnAlaPheValValProGlnGlnAlaVal---ThrArgGlyAlaLysAspThrValMet 329

Db 1036 GCGCGCACCTGATCAAGACGACGAGCGCTGCGACCGACCTGCGCAAGAGTTCGTGCTG 1095

QY 330 IleValAsnAlaGlnGlyLysMetGluProArgGlnValThrValAlaGlnGlnGlnGly 349

Db 1096 GTCTCGAATGGCGACAAAGACCTGTCACCGCTGAGAGTGGACCGCAAGCTGCGAG 1155

QY 350 ThrAsnTrpIleValThrSerGlyLeuLysAspGlyAspLysValValValGlnGlyIle 369

Db 1156 GCGCTGCGATCGTCCGCGCGCGCTGACAGCGCGCGCGCGCGCGCGCGCGCGCGCTG 1215

QY 370 SerIleAlaGlyIleThrGlyAlaLysLysValThrProLys-----GluTrpAlaSer 387

Db 1216 CAG-----CGGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1269

QY 388 SerGluAsnGlnAla-----AlaAlaProGlnSerGlyValGlnThrAlaSerGluAla 405

Db 1270 GCCGACACCTGCGCACCTCGCGCGCGCTGCGCGAGTCTGCGCGACGACGACGACCG 1329

QY 406 LysThrAlaSerGluAlaGlu 412

Db 1330 AAGGTGGCGGCTCCCAAGAC 1350

Db 1330 AAGGTGGCGGCTCCCAAGAC 1350

Db 1330 AAGGTGGCGGCTCCCAAGAC 1350

Db 1330 AAGGTGGCGGCTCCCAAGAC 1350

Db 1330 AAGGTGGCGGCTCCCAAGAC 1350

Db 1330 AAGGTGGCGGCTCCCAAGAC 1350

Db 1330 AAGGTGGCGGCTCCCAAGAC 1350

Db 1330 AAGGTGGCGGCTCCCAAGAC 1350

Db 1330 AAGGTGGCGGCTCCCAAGAC 1350

Db 1330 AAGGTGGCGGCTCCCAAGAC 1350

Db 1330 AAGGTGGCGGCTCCCAAGAC 1350

Db 1330 AAGGTGGCGGCTCCCAAGAC 1350

Db 1330 AAGGTGGCGGCTCCCAAGAC 1350

Db 1330 AAGGTGGCGGCTCCCAAGAC 1350

Db 1330 AAGGTGGCGGCTCCCAAGAC 1350

Db 1330 AAGGTGGCGGCTCCCAAGAC 1350

Db 1330 AAGGTGGCGGCTCCCAAGAC 1350

Db 1330 AAGGTGGCGGCTCCCAAGAC 1350

Db 1330 AAGGTGGCGGCTCCCAAGAC 1350

Db 1330 AAGGTGGCGGCTCCCAAGAC 1350

Db 1330 AAGGTGGCGGCTCCCAAGAC 1350

Db 408 CGCTTCGAGAACCGCGACGAGCTGTTCCTCCCAACGATTCGTCACGTCGCTGCTG 349
Qy 306 GlnValAlaValAspAsnAlaPheValAlProGlnGlnAlaValThrArgGlyAlaLys 325
Db 348 GCGCATACCTTCAGAGCGGTGCTGACATTCACGCAACGCGTGCAGCGGACCGAAC 289
Qy 326 AspThr---ValMetIleValAsnAlaGlnGlyMetGluProArgGluValThrVal 344
Db 288 GGTATCTATGTGTCGTGTGTCGCGCCGACAAAGATGTCAGCAGCGAGGTGCGCATC 229
Qy 345 AlaGlnGlnGlnGlyThrAsnTrpIleValThrSerGlyLeuLysAspGlyAspLysVal 364
Db 228 GGCACACGACGAGAACGACGCGGTGTGTGTGAAGCGCTGAAAGCCGCGGACGAGGTG 169
Qy 365 ValValGlnGly 368
Db 168 GTGGTGAAGGC 157
RESULT 17
US-09-252-991A-7520/c
; Sequence 7520, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 7520
; LENGTH: 1386
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7520

Alignment Scores:
Pred. No.: 1,14e-33 Length: 1386
Score: 396.50 Matches: 126
Percent Similarity: 47.55% Conservative: 58
Best Local Similarity: 32.56% Mismatches: 171
Query Match: 19.64% Indels: 32
Gaps: 7

US-09-889-756a-2 (1-412) x US-09-252-991A-7520 (1-1386)
Qy 28 GlnGlyAspAlaValGlnGlyGlnProAlaArgGluAlaProAlaProVal--- 46
Db 1236 GCGCGTGGGATGCTGCTGCGGCGGACCGCGGAGGCGCGCGCGCGCGCGCGCGCG 1177
Qy 47 -----ValGlyValAlaThrValHisProGlnThrValAlaLeuThrValGluLeuPro 64
Db 1176 GTCCCGGTGAACGTGGCGCGGTGGAGGCGGCGACGTCGACGACAAAGATGAGGCGCATC 1117
Qy 65 GlyArgLeuGlnSerLeuArgThrAlaAspValArgAlaGlnValGlyIleIleGln 84
Db 1116 GGCACGCGGACTTGTGTCACACGCTGTATCCGACCCAGATCGACGCGAGTTCGCC 1057
Qy 85 LysArgLeuPheGlnGlnGlySerTyrValArgAlaGlnGlnProLeuTyrGlnIleAsp 104
Db 1056 GCGCTGCTGTGAGCGGAGGCGCAATGTGTCGAGGCGGCGGCTGTGCGACCAATGCAT 997
Qy 105 SerSerThrTyrGluAlaAsnLeuGlnSerAlaArgAlaGlnLeuAlaThrAlaGlnAla 124
Db 996 GACCGCGCGCTGTGCTGCGCGCTGTGAGCGAGCTCAGGCTCCAGGCGAGCAACGAGCC 937
Qy 125 ThrLeuAlaLysAlaAspAlaAspLeuAlaArgTyrLysProLeuValAlaAlaGlnAla 144
Db 936 CAGCTGAATCCGCGGACGAGGACCTGTCACGCTATGCGGACGCGTGC 877

Qy 145 ValSerArgGlnGlnIleTyrAspAlaAlaValThrAlaLysArgSerAlaGlnAlaGlyVal 164
Db 876 GTTCGCGGCGCACTGCTGAGACGACGACGACGACGACGACGACGACGACGACGACGAC 817
Qy 165 LysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsnArgSerArgIleThr 184
Db 816 AAGGCCAAGCATGTCACCATTCACGCGGAGCGGAGCGGCTGTCTTACCCCGATCAC 757
Qy 185 AlaProIleSerGlyPheIleGlyGlnSerLysValSerGluGlyThrLeuLeuAla 204
Db 756 TCGCGGATTCGCGCAAGTCGTGTATCCGCAACCTTCATGTGCGCAACTGTGCGGATC 697
Qy 205 GlyAspThrThrValIleAlaThrIleArgGlnThrAsnProMetTyrValAsnValThr 224
Db 696 GCGCACAGTCTCGGCGCTGTTCAGGTCACCAAGATCCGACGATCTCCGATCTTCTCC 637
Qy 225 GlnSerAlaSerGluValMetLysLeuArgArgGlnIle-----AlaGlnGlyLys 241
Db 636 CTGCACAGGACAGTTCGCTCCAGATTCAGGCGCTGCGCGGCGGACGCGCGCTGCGC 577
Qy 242 LeuLeuAlaAlaAspGlyValIleAlaValGlyIleLysPheAspArgIleThrValTyr 261
Db 576 GCTTACAGCGCGACGCGCGGAGCGCTGCGC----- 544
Qy 262 ProGluArgGlyArgLeuLeuPheAlaAspProValValAsnGlnSerThrGlyGlnIle 281
Db 543 -----GAGGCGCGGTGCTGACCATGACACACCATGACCATGCTCCAGTCACATC 490
Qy 282 ThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetProGlyLeuTyrValArg 301
Db 489 CGCGTGGCGCTCTTCATTAACCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCG 430
Qy 302 ValLeuMetAspGlnValAlaValAspAsnAlaPheValAlProGlnGlnAlaValThr 321
Db 429 GTGAGCTTGACACACCGGCGTCCGCGCGACGACTGTGCTGTGACCAAGCGGTGCGC 370
Qy 322 ArgGlyAlaLysAspThrValMetIleValAsnAlaGlnGlyMetGluProArgGlu 341
Db 369 GCGCGCTGAGAGGCAATTCCTGTCATCCGGGTGCGCGCGCGCGCGCGCGCGCGCGCA 310
Qy 342 ValThrValAlaGlnGln---GlnGlyThrAsnTrpIleValThrSerGlyLeuLysAsp 360
Db 309 GTGCGGCTGTCCGACGATCGACGCGCTCAGT-----GTGGTGAAGCGCTGCGCAGC 256
Qy 361 GlyAspLysValValAlaGlnGlyIleSerIleAlaGlyIleThrGlyAlaLysVal 380
Db 255 GGTGACGAGGTGTGTGACGCGGCGCTCG-----CGGCTG 220
Qy 381 -ThrProLysGluTrpLaseSerGluAsnGlnAlaAla-----ProGlnSe 397
Db 219 ATGCGCGCGCGCTGTGTCATTCAGAGCCGCGCGCGCGCGCGCGCGCGCGCGGAG 160
Qy 397 GlnValGlnThrLase 403
Db 159 CGGACGCGGTGACCGCGCG 141
RESULT 18
US-09-252-991A-7817
; Sequence 7817, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142


```

CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 1086
LENGTH: 4404
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1086

Alignment Scores:
Pred. No.: 2,41e-28 Length: 4404
Score: 355.00 Matches: 105
Percent Similarity: 49.73% Conservative: 78
Best Local Similarity: 28.53% Mismatches: 168
Query Match: 17.58% Indels: 18
DB: Gaps: 6

US-09-889-756a-2 (1-412) x US-09-252-991A-1086 (1-4404)

Qy 18 ValAlaLeuValLeuSerSerCysGlyLysGlyValAspAlaAlaGlnGlyGlnPro 37
Db ||||| : : : : : ||||| : : : : :
Qy 131 GTCCGGCCCTTATCGCTGCTGCGGGAATCGCGC-----CCG 169
Db ||||| : : : : : ||||| : : : : :
Qy 38 AlaGlyArgGlnAlaProAlaProValValGlyValThrValHisProGlnThrVal 57
Db ||||| : : : : : ||||| : : : : :
Qy 170 CCGGGGGCGGGAGCCGCCCGCCAGCGTTCCCGCCAGGTGGTGGTCCCGCGT 229
Db ||||| : : : : : ||||| : : : : :
Qy 58 AlaLeuThrValGluLeuProGlyArgLeuGlnSerLeuArgThrAlaAspValArgAla 77
Db ||||| : : : : : ||||| : : : : :
Qy 230 ACACCTACGCGCGATTACCGGCTCGCTGACCGCTGACGAGTAGAATCGCGCG 289
Db ||||| : : : : : ||||| : : : : :
Qy 78 GlnValGlyGlyLeuGlnGlnLysArgLeuPheGlnGlnGlySerTyValArgAlaGly 97
Db ||||| : : : : : ||||| : : : : :
Qy 290 CGCGTCCCGCGCTACATCCAGAGCGTTACGCTGCGGAGGCGCCCTGTAGAGAGAGCGC 349
Db ||||| : : : : : ||||| : : : : :
Qy 98 GlnProLeuTyArgGlnLeuAspSerSerThrTyGlnAlaLeuGlnSerAlaArgAla 117
Db ||||| : : : : : ||||| : : : : :
Qy 350 CAGAACTGTTCTTCATCGACCCCGCGGTTCAGCGGGCGGAGATGCCCGCCAGGCA 409
Db ||||| : : : : : ||||| : : : : :
Qy 118 GlnLeuAlaThrAlaGlnAlaThrLeuAlaLysAlaAspAlaAspLeuAlaArgTyLys 137
Db ||||| : : : : : ||||| : : : : :
Qy 410 CGCTCGCGGAGCGCGAGCGCGCTGCTGCGGCGACCGAACAACGAGCGCGAG 469
Db ||||| : : : : : ||||| : : : : :
Qy 138 ProLeuValAlaAlaGlnAlaValSerArgGlnGlnTyArgAlaAlaValThrAlaLys 157
Db ||||| : : : : : ||||| : : : : :
Qy 470 CTGCTGTATGCGCGAGGTCGTGCTGCGGAGGCGGCTGACAGCGCATCGCTCGCGC 529
Db ||||| : : : : : ||||| : : : : :
Qy 158 ArgSerAlaGlnAlaGlyValLysAlaAlaGlnAlaAlaLysSerAlaGlyLys 177
Db ||||| : : : : : ||||| : : : : :
Qy 530 AATGCCAGAGGCGCGAGTCGATCGAGGCGCGCCCTCGACGCGGCGCAACTGAT 589
Db ||||| : : : : : ||||| : : : : :
Qy 178 LeuAsnArgSerArgLysAlaAlaProLysSerArgPheLysGlnSerLysValSer 197
Db ||||| : : : : : ||||| : : : : :
Qy 590 CTGCGCTTACGCGGCGTACGCGACCGATCGCGGCGGTGCGGCGCATTCAGTACC 649
Db ||||| : : : : : ||||| : : : : :
Qy 198 GlnGlyThrLeuLeuAlaGlyAspThrThrValLeuAlaThrLysGlnThrAsn 217
Db ||||| : : : : : ||||| : : : : :
Qy 650 GAGGGCAACTACGACCAATGCG---GTCAACCGCGCTGACCAAGCATCGTTCCGTGAT 706
Db ||||| : : : : : ||||| : : : : :
Qy 218 ProMetTyVal---AsnValThrGlnSerAlaSerGlnValMetLysLeuArgGln 237
Db ||||| : : : : : ||||| : : : : :
Qy 707 CCGCTAATCGTGAATCTTCGATGTCGACGAGCGCACTACCTGACGCGCCCTGCGGCGCAGC 766
Db ||||| : : : : : ||||| : : : : :
Qy 237 AlaGlnGlyLysLeuLeuAlaAlaAspGlyValLysAlaAlaValGlyLysPheAspAs 257
Db ||||| : : : : : ||||| : : : : :
Qy 767 CGCGGAGGAGGAGCGAG---CAGGCGCCGAGGTC---AAGTGGCGTGTCAACCGACGA 822
Db ||||| : : : : : ||||| : : : : :
Qy 257 pGlyThrValTyProGlnLysGlyValGlnLeuPheAlaAspProValValAsnGln 277
Db ||||| : : : : : ||||| : : : : :
Qy 823 GTCC-----TATGAGCAAGCAAGTCGCTGATTCCTCGCAACCGCCGACCGCGG 876

```

```

Qy 277 rThnGlyGlnLysLeuLeuAlaAlaValProAsnAspGlnAsnLysLeuMetProG 297
Db ||||| : : : : : ||||| : : : : :
Qy 877 GACCGGACGAGTCCGGGTTCGGGGGTGTGACAAACCCGAGCGGCACTGACGCGCGG 936
Db ||||| : : : : : ||||| : : : : :
Qy 297 yLeuTyValArgValLeuMetAspGlnValAlaValAspAsnAlaPheValValProG 317
Db ||||| : : : : : ||||| : : : : :
Qy 937 GCTGTTCGCAAGGTCTGAGCGGCAAGCCCTGCGCGGAGTGTGCTGCGCGCA 996
Db ||||| : : : : : ||||| : : : : :
Qy 317 nGlnAlaVal---ThrArgLysAlaLysAspThrValMetLysValAsnAlaGlnGly 336
Db ||||| : : : : : ||||| : : : : :
Qy 997 CCATTCCATGCGACGACCAAGCGCGCTATGTGCTGATGTCGACGAGGCAACAA 1056
Db ||||| : : : : : ||||| : : : : :
Qy 336 yMetGlnProArgGlnValThrValAlaGlnGlnGlnTyThrAsnTrpLysValThr 356
Db ||||| : : : : : ||||| : : : : :
Qy 1057 GACCCAGTACCGCGCGGTGAGACTCGGCGCGGATGTGACAGCGCTGCGGCGTCCGCCA 1116
Db ||||| : : : : : ||||| : : : : :
Qy 356 rGlyLeuLysAspGlyAspLysValValValGlnGlyLysSerLysAlaGlyLysThr 376
Db ||||| : : : : : ||||| : : : : :
Qy 1117 GGGCTTCAGCCGCGGCGGAGCGCATCGTCTCAAGGCGCTG-----GTCCGCGC 1164
Db ||||| : : : : : ||||| : : : : :
Qy 376 yAlaLysLysValThrProLys 383
Db ||||| : : : : : ||||| : : : : :
Qy 1165 GGAATGCAATGATCAACCGCGC 1186
Db ||||| : : : : : ||||| : : : : :

RESULT 22
US-09-252-991A-14946
Sequence 14946 Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 14946
LENGTH: 1224
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14946

Alignment Scores:
Pred. No.: 1.22e-27 Length: 1224
Score: 340.50 Matches: 104
Percent Similarity: 51.52% Conservative: 66
Best Local Similarity: 31.52% Mismatches: 145
Query Match: 16.86% Indels: 15
DB: Gaps: 8

US-09-889-756a-2 (1-412) x US-09-252-991A-14946 (1-1224)

Qy 91 GlySerTyValArgAlaGlyGlnProLeuTyArgGlnLysAspSerSerThrTyGlnAla 110
Db ||||| : : : : : ||||| : : : : :
Qy 2 GGGGCACTGTGTAAGAAAGCGCACTGCTGTTCCATGCAACCGCGCGCTGCGAGGCC 61
Db ||||| : : : : : ||||| : : : : :
Qy 111 AsnLeuGlnSerAlaArgAlaGlnLysAlaThrAlaGlnAlaThrLeuAlaLysAla 130
Db ||||| : : : : : ||||| : : : : :
Qy 62 GAGGTCAAGCGCTCGAAGCCGAGGTGCAACAGGCGCGCGCGCCAGGCGCGAGGCTC 121
Db ||||| : : : : : ||||| : : : : :
Qy 131 AlaAspLeuAlaArgTyLysProLeuValAlaAlaGlnAlaValSerArgGlnGln 150
Db ||||| : : : : : ||||| : : : : :
Qy 122 AACGAAGCCAGGCGCGCAAGCGCTGCGCGGCGGACGCGATCTCCGGGAATTCGCC 181
Db ||||| : : : : : ||||| : : : : :
Qy 151 AspAlaAlaValThrAlaLysArgSerAlaGlnAlaGlyValLysAlaAlaGlnAla 170
Db ||||| : : : : : ||||| : : : : :
Qy 182 GACGCGCGACCAACCGCCCGCAGGAAGCCAGAGCGCGGTGCGCGCGACCCGAGCGCA 241

```

```

Qy 171 ILeysSerAlaGlyIleAsnLeuAsnArgSerArgIleThrAlaProIleSerGlyPhe 190
Db 242 CTGAGCGGGCGCGCGCTGAACCTGAGCTTCAACCGGATCAACCGCGCATGACGAGTCC 301
Qy 191 ILeGlyGlnSerIleValSerGlyGlyThrLeuLeuAsnAlaGlyAspThrThrValLeu 210
Db 302 GTCAGCGCGCGCGGATCAACCGCGGCACTGGTCACTCGGAGAA--ACCTGCTC 358
Qy 211 AlaThrIleArgGlnThrAsnProMetCysValAsnValThrGlnSerAlaSerGluVal 230
Db 359 ACCACCTGCTGACACCGACCAAGGTCTACCTCACTTCCAGCGCGCAACGAGCGGTGTC 418
Qy 231 MetIys-----LeuArgArgGlnIleAlaGlyIleValLeuLeuAlaAlaAspGly 247
Db 419 CTCAGATGACGTGCGAGCTGGCCCGCAG-----GCCGGTCCGACACCGCGCAGAGAGC 472
Qy 248 ValIleAlaValAlaGlyIleValPheAspAspGlyThrValIleProGlyIleValArgLeu 267
Db 473 CCGGTCTACCTGCGCTGAGCAGCGAGAGCGCAAC-----CCGCACTGGGCGCGCTG 526
Qy 268 LeuPheAlaAspProValValAlaGlnIleThrGlyGlnIleThrLeuArgAlaAlaVal 287
Db 527 GACTTCTCCGACCAACAGGTCAACCGCGCTACCGGACCATCCGCGCGCGCGCTGTC 586
Qy 288 ProAsnAspGlnAsnIleLeuMetProGlyLeuValArgValLeuMetAspGlnVal 307
Db 587 GACAACGCGCAAGGGGAGATTCAACCGCGCTCAACGCGCTGAGAGTGTGCGGAGC 646
Qy 308 AlaValAspAsnAlaPheValValProGlnIleAlaVal---ThrArgGlyAlaValAsp 326
Db 647 AAGACTACGCGCGCACCTCATCAGAGCAGAGCGGCGGACCGACCTGGGCGCAAGAG 706
Qy 327 ThrValMetIleValAlaGlnIleGlyIleValProArgGluValThrValAlaGln 346
Db 707 TTCTGCTGCTGCTGCTGATGCGCAGACCAACCGCTTACCGCACCGTCAAGATGGAGCC 766
Qy 347 GlnGlnIleThrAsnTrpIleValIleThrSerGlyLeuValAspGlyAspValValVal 366
Db 767 AAGTGGAGGGCGCTGCGCATCTGCGCAGCGCGCTGACCAAGGGGAGCGGATGCTGTG 826
Qy 367 GlnGlyIleSerIleAlaGlyIleThrGlyAlaValValIleThrProIys-----Glu 384
Db 827 AATGCGCTGCAG-----CGGGTCCGCGCGCATGAGTGCAGTCCGACAGAGTCCAG 880
Qy 385 TrpAlaSerSerGluAsnGlnAla-----AlaAlaProGlnSerGlyValAlaIleThrAla 402
Db 881 ATGGCCAGCGCGCAGACCCCTGACCCCTCGCGCGCTGCGGCACTGCTGGCGACAGC 940
Qy 403 SerGluAlaValThrAlaSerGluAlaGlu 412
Db 941 GAACCAACGAGGTGCGCGCTCCAGAGAC 970

RESULT 23
US-09-557-884-1
Sequence 1, Application US/09557884
Patent No. 6506581
GENERAL INFORMATION:
APPLICANT: Fleischmann et al.
TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSER: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22

```

```

SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/557,884
FILING DATE: 25-Apr-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,102
FILING DATE: JUN-5-1995
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: P8186P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Alignment Scores:
Pred. No.: 1,13e-16 Length: 1830121
Score: 286.00 Matches: 105
Percent Similarity: 41.01% Conservative: 66
Best Local Similarity: 25.18% Mismatches: 148
Query Match: 14.17% Indels: 98
DB: Gaps: 14

US-09-889-756a-2 (1-412) x US-09-557-884-1 (1-1830121)
Qy 6 PheIysAlaMetArgAlaAlaIleuAlaAlaValAlaIleuValIleuSerSerCys 25
Db 946822 TTTAATATGATTAAGGCGTATGATGAGCGAGCCATTGCA----- 946863
Qy 26 GlyIysGlyAspAlaAlaGlnIleGlyIleProAlaGlyArgGluAlaProAlaPro 45
Db 946864 -----GGAATGCCA-----GAACTTCAAGCCCA 946887
Qy 46 ValValIleValValIleThrValHisProGlnThrValAlaIleuThrValGluLeuProGly 65
Db 946888 GTAACC---GCACCTGAAGTTCAACCGCGTGAATGAGCCGACGATTATTAACAAACAGT 946944
Qy 66 ArgLeuGlnSerLeuAsnGlyThrAlaAspValArgAlaGlnValAlaGlyIleIleGlnIys 85
Db 946945 CTGTGGCTCCAAATCAAGGCGCATGCTCAGTACACAAATGCGGCGCGCTTCACAA 947004
Qy 86 ArgLeuPheGlnIleGlySerIleValArgAlaGlyGlnProLeuValIleAspSer 105
Db 947005 GTACTTGTCAAAATGACCAAAATGTGAAGGAGTACGCTGTGAGACCTTGATAGT 947064
Qy 106 SerThrTyGlnAlaAsnLeuGlnSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThr 125
Db 947065 TCTGTGAAGAGCTATCTACAAAGTGTGACGACCAATTATCAGACACTTCGCAACT 947124
Qy 126 LeuAlaIysAlaAspAlaAspLeuAlaArgTyIleProLeuValAlaAlaGluAlaVal 145
Db 947125 -----TACCAACGTTATGTGGTTATTAATTAACCAATGCTGA 947163
Qy 146 SerArgGlnIleValAspAlaAlaValThrAlaIysArgSerAlaGluAlaGlyValIys 165
Db 947164 TCACGTCAAGAAATGATTAACGCAAAAGCGCTTATATGCTCAAGTACGATGATTTGAA 947223
Qy 166 AlaAlaGlnAlaAlaIleIysSerAlaGlyIleAsnLeuAsnArgSerArgIleThrAla 185
Db 947224 TCTCTAAAGCAGCAATTGAA-----CGTCGTAATATGTTGCG 947262
Qy 186 ProIleSerGlyPheIleGlyGlnSerIleValSerGlyIleThrLeuLeuSerAlaGly 205
Db 947263 CCATTGATGCAAGCAGAGTATGTGAAATCAATGTGCAATATGTGAATGTGGA 947322

```

QY 206 AspThrValLeuAlaThrIleArgGlnThrAsnProMetYrValAsnValThrGln 225
Db 947223 -----ACAGAAATGTGCGGTAGAGATCTAGCTCAATGAAGTGGATTGCTCTT 947376
QY 226 SerIleSerGluValMetLysLeuArgArgGlnIleAlaGluGlyLysLeuLeuAla 245
Db 947377 TCACAAATGATTTAGATAAATTCAT-----ATCGGTCAGCGGTTACAGCG 947424
QY 246 AspGlyValIleAlaValGlyLysLeuAspAspGlyThrValTyProGluGly 265
Db 947425 ACAACAGATGCTCGCTGGGGAACATTT-----TCAGCT 947460
QY 266 ArgLeuPheLeuAlaAspProValValAsnGluSerThrGlyGlnIleThrLeuArgAla 285
Db 947461 CGATCACTGCGGATTAACCGCATTAATTCATCAACAGCTTATAGTATGTTACAGCT 947520
QY 286 AlaVal---ProAsnAspGlnAsnIleLeuMetProGlyLeuTyrValArgValLeuMet 304
Db 947521 ACTTTGATCTGGAAGATGGCATTAATTCCTTCAGGTATGTTCTCTCGCTTACGAT 947580
QY 305 AspGlnValAlaValAspAsnAlaPheValValProGlnGlnAlaValThr----- 321
Db 947581 GCACCTTCCAACTGAACAAATCAAGTTGTCGTCACAGTATGCTTACAAATATG 947640
QY 321 ----- 321
Db 947641 TATGGCGAATTCCTATTACTTGAACCATTCATCGAAGAAAGAAAGAAATATGCA 947700
QY 322 -----ArgGlyAlaLysAspThrValMetIleValAsn 332
Db 947701 GGTAATGAATAATGATCGCTCTATCGTCGGAACAGATCAACGATTAATTAAGAT 947760
QY 333 AlaGlnGlyMetGluProArgGluValThrValAlaGlnGlnGlnGlyThrAsnTrp 352
Db 947761 CGTAAAGT-----GTTATATGCTCAATTACAGGGAATGAA--- 947796
QY 353 IleValThrSerGlyLeuLysAspGlyAspLysValValAlaGluGlyIleSerIleAla 372
Db 947797 -----GTTAAAGTGGAGATTAATTAATTAACAGCGGTACAGAA----- 947835
QY 373 GlyIleThrGlyAlaLysValThrProLysGluTrpAlaSerSerGlu 389
Db 947836 GGATATGTAATGAAGCTGTGT-----GAATGATTAATAAAGAC 947877

RESULT 24
US-09-643-990A-1
Sequence 1, Application US/09643990A
Patent No. 6528289
GENERAL INFORMATION:
APPLICANT: Robert D. Fleischmann
Mark D. Adams
Owen White
Hamilton O. Smith
J. Craig Venter
TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
thereof, and Uses Thereof
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville,
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/643,990A

FILING DATE: 23-Aug-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,429
FILING DATE: 1995-06-07
APPLICATION NUMBER: 08/426,787
FILING DATE: 1995-04-21
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB186P1C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-610-5790
TELEFAX: 310-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Alignment Scores:
Pred. No.: 1,13e-16 Length: 1830121
Score: 286.00 Matches: 105
Percent Similarity: 41.01% Conservative: 66
Best Local Similarity: 25.18% Mismatches: 148
Query Match: 14.17% Indels: 98
Gaps: 14

US-09-889-756a-2 (1-412) x US-09-643-990A-1 (1-1830121)

QY 6 PheLysAlaMetArgAlaAlaLeuAlaAlaValAlaLeuValLeuSerSerCys 25
Db 946822 TTTAATATGATTAAGGCGTAATGATTAAGCGGCACTTCA----- 946863
QY 26 GlyLysGlyLysPheAlaAlaGlnGlyGlnProAlaGlyArgGluAlaProAlaPro 45
Db 946864 -----GAAATGCCA-----GAAATCTCAAGCCCA 946887
QY 46 ValValGlyValValThrValHisProGlnThrValAlaLeuThrValGluLeuProGly 65
Db 946888 GTAACC---GCACCTGAAGTTCAACCGCGTAATGACGCGCATTAACACAACAGGT 946944
QY 66 ArgLeuGluSerLeuArgThrAlaAspValArgAlaGlnValGlyIleIleGlnLys 85
Db 946945 CTGTGCGTCAATCAAGCGCATGCTCAGTACACAAATGCGCGCTTCACAA 947004
QY 86 ArgLeuPheGlnGluGlySerTyValArgAlaGlyGlnProLeuTyArgIleAspSer 105
Db 947005 GTACTTGTCAAAATGACAAATATGTAAGGAGGAGGCTTGTGAGAGCTGATAGT 947064
QY 106 SerThrTyArgLysLeuLeuGluSerValArgAlaGlnLeuAlaThrAlaGlnAlaThr 125
Db 947065 TCGTTGAACAGCTATCTACAAGCTGCGCAGGCAAAATTAACGACCTTCGTAACACT 947124
QY 126 LeuAlaLysAlaAspAlaAspLeuAlaArgTyLysProLeuValAlaAlaGluAlaVal 145
Db 947125 -----TACCAACGTTATGCTGCTTATTAATTAACAAAGCTGTA 947163
QY 146 SerArgGlnGluTyArgPheAlaAlaValThrAlaLysArgSerAlaGluAlaGlyValLys 165
Db 947164 TCACGTCAAGAAATGGAATTAACGAAAGCGGCTTATGATCTCAAGTACAGTATGAA 947223
QY 166 AlaAlaGlnAlaAlaIleLysSerAlaGlyLysLeuAsnArgSerArgIleThrAla 185
Db 947224 TCTCTAAACACACCAATGAA-----CGTCAATAATGTTGCG 947262
QY 186 ProIleSerGlyPheIleGlyGlnSerLysValSerGluGlyThrLeuLeuAlaGly 205
Db 947263 CCATTTGATGCAAGACGATTAATGTAATCAATGTTGACAAATATGTAATGTTGGA 947322

```

OY 206 AsphtrThrValLeuAlaThrIleArgGlnThrAsnProMetTyrValAsnValThrGln 225
Db 947323 -----ACAGAAATGTGCGTGTGAGATGACTACTGATCAATGAAGATGATTTGCTCTT 947376
OY 226 SerIAspGlnValMetLeuLeuArgArgGlnIleAlaGlnGlyLeuLeuAlaIle 245
Db 947377 TCACAAATGATTGATTAATTAATTCAT-----ATCGGTACACGGCGGTTCACACGG 947424
OY 246 AspGlyValIleAlaValGlyIleLeuSphaeSparGlyThrValTyrProGlnLeuSgly 265
Db 947425 ACAACAGATGCTCGCTTGGCGCAACATTT-----TCACCT 947460
OY 266 ArgLeuLeuPheAlaAspProValValAsnGlnSerThrGlyGlnIleThrLeuArgAla 285
Db 947461 CGAATCACTGCGATTAACCTGCGCATTAATCATCAACAGGTTTGTGATGTTTCAGGCT 947520
OY 286 AlaVal-----ProAsnAspGlnAsnIleLeuMetProGlyLeuTyrValArgValLeuMet 304
Db 947521 ACTTTGATCCTGGAAGATGGCATTAATGCTTTCAGGATGTTCTCTCGCTTACGCAAT 947580
OY 305 AspGlnValAlaValAspAsnAlaPheValValProGlnGlnAlaValThr----- 321
Db 947581 GCACCTCCACAGCAACAAATCAAGTTGCTGTCACAGTAGATTAAGCTACAAATATG 947640
OY 321 ----- 321
Db 947641 TATGCGAAATGCTCTATTACTTGAAACCTATATCTGAAGAAAGAAAGAAATGTCATCA 947700
OY 322 -----ArgGlyAlaValSparThrValMetIleValAsn 332
Db 947701 GGTAAATGAAATTTGATGCTGCTCTATGTCGAAACAGATCAACGTAATTTACTTAAGAT 947760
OY 333 AlaGlnGlyGlyMetGluProArgGlyValThrValAlaGlnGlnGlnGlyThrAsnTyr 352
Db 947761 CGTCAAGCT-----GTTATGCTCAATTACAGGAAATGAA----- 947796
OY 353 IleValThrSerGlyLeuLysAspGlyAspLysValValValGlyGlyIleSerIleAla 372
Db 947797 -----GTTAAAGTGGAGATTAATAATTATTAACAGCGGTCACGAA----- 947835
OY 373 GlyIleThrGlyAlaValLysValThrProLysGlnTyrPalasSerSerGlu 389
Db 947836 GGTATGTGTAATGGAAGCTTTGTG-----GAATGATTAATAAAGAC 947877

RESULT 25
US-09-252-991A-7914
; Sequence 7914, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 7914
; LENGTH: 1185
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7914

Alignment Scores:
Pred. No.: 9,94e-21 Length: 1185
Score: 277.00 Matches: 104
Percent Similarity: 45.57% Conservative: 71
Best Local Similarity: 27.08% Mismatches: 141
Query Match: 13.72% Indels: 68
DB: 4 Gaps: 14

```

```

US-09-889-756A-2 (1-412) x US-09-252-991A-7914 (1-1185)
OY 14 LeuAlaIleAlaValAlaLeuValLeuSerSerGlyGly-----LysGlyGlyAspAlaIle 32
Db 82 CTCGCCGGGCGGATGCGCGGTGGTGGCATTTCTCCCGGCTTACAGAGTCTACTCATTCCT 141
OY 33 GlnGlyGlnProAlaGlyArgGluAlaProAlaProValValGlyValThrVal 52
Db 142 CAG-----CAGATCGCCTTTTCAGGCCACCGAAACCGCGCATCAGCGTG----- 186
OY 53 HisProGlnThrValAlaLeuThrValGluLeuPro----- 64
Db 187 -----ACCGCACCTCGCGCGAAAGCGTCTCGGAGAGCCGCTCCGACCATC 237
OY 65 GlyArgLeuLeuSerLeuArgThrAlaAspValArgAlaGlnValGlyIleIleGln 84
Db 238 GGCAGCCTCAAGGATTCACAGGGCGGTGACCTTCAACCGCGAAGTCTCCGACCGTACGC 297
OY 85 LysArgLeuPheGlnGlnGlySerTyrValArgAlaGlyGlnProLeuTyrGlnIleAsp 104
Db 298 GACGTACTGTCTTCCGCGGACCAAGTGAAGCTGACCAACCGCTGATCCAGTGGAA 357
OY 105 SerSerThrTyrGluAlaAsnLeuGlnSerAlaArgAlaGlnLeuAlaThrAlaGlnAla 124
Db 358 AGCAGCGTCAGAGAGCCACCTGCGC-----ACTGCGAGGCG 396
OY 125 ThrLeuAlaLysAlaAspAlaAspLeuAlaArgTyrLysProLeuValAlaAlaGluAla 144
Db 397 GATCTCGCGCTGCGCGAGCGCGATTCACAGCGCGCGCGCAACTGATCGGACGAGAGCC 456
OY 145 ValSerArgGlnGlyTyrAsp-----AlaAlaValThrAlaLysArgSerAlaGluAla 162
Db 457 ATCTGAAAGCGAATTCGATGCTGCTGCGCGCGAGTGGCCAAAGACAGCGGACACCGTC 516
OY 163 GlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsnArgSerArg 182
Db 517 -----GCCGAGCTGAAAGCGCGC-----CTGGGAAAGAAACGCG 549
OY 183 IleThrAlaProIleSerGlyPheIleGlyGlnSerTyrValSerGlyGlnThrLeu 202
Db 550 GTGCTCGCGCTTCCGCGGACCATGCGATCCGCGAGGTGACGTGCGGACTACGAT 609
OY 203 AsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAsnProMetTyrValAsn 222
Db 610 TCGCGCGCG-----ACGCCGATCGCACCTTGACAGCACTTTCACACCTGCTCTGGAT 663
OY 223 ValThrGlnSerAlaSerGluValMetLysLeuArgArg--GlnIleAlaGlnGlyLys 241
Db 664 TTCACCTGCGCGAGCAGCACTTCCCTGCTCAGCGCGCGGACGCTGTGAAGTCCGG 723
OY 242 LeuLeuAlaIleAspGlyValIleAlaValGlyIleLysPheAspArgGlyThrValTyr 261
Db 724 GTGCGCGCTTACCGCGCGCGAGTG-----TTCCAGC----- 753
OY 262 ProGlnLysGlyArgLeuLeuPheAlaAspProValValAsnGlnSerThrGlyGlnIle 281
Db 754 -----GCCAGATGCGCGCGCATCAACCCAGGTGACAAACGAGACCGGAACTCG 804
OY 282 ThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetProGlyLeuTyrVal-- 300
Db 805 CAGGTCCGCGCTGCTCGGAGAACCCGAGCGCAAGCTGCTGCGGCGCATGTCGCCAAC 864
OY 301 -----ArgValLeuMetAspGlnValAlaVal-- 309
Db 865 CTCGAGGTGATGTTGCTGCGCGAGAACAAACGCTGCTGCGCGGAGCGGCGCATACC 924
OY 310 -----AspAsnAlaPheValValProGlnGlnAlaValThrArgGlyAla 324
Db 925 TTCACCTTACGCGGACGATGATCTGCTGCTGCGGAGAAAGAAAGACGAGCGGCGAG 984
OY 325 LysAspThrValMetIleValAsnAlaGlnGlyGlyMetGluProArgGlyValThrVal 344
Db 325 ----- 344

```


DB: 4 Gaps: 11

US-09-889-756a-2 (1-412) x US-09-328-352-3771 (1-1359)

QY 65 G|YArgLeuG|uSerLeuArgThrAlaAspValAlaArgAlaGlnGlyIleLeuGln 84

DB 178 GGTACACTGTGATGACCAAGCTTATTAAGTGGGTGCTGAGTATCGTCAGATTAA 237

QY 85 LysArgLeuPheGlnG|uSerLeuArgThrAlaArgAlaGlnGlyIleLeuGln 104

DB 238 AAGATGATGTGACGCTTGTGATCAAGTAAACAGCTCACTTATTCGCAAAATGGAC 297

QY 105 SerSerThrTyrg|u-----AlaAsnLeuG|uSerAlaArgAla 117

DB 298 TCGACCAACCAAGAAACAGTTAAACATCTGATGATATTAATAATTGAGGCA 357

QY 118 GlnLeuAlaThrAlaGlnAlaThrLeuAlaLysAlaAspLeuAlaArgTyrls 137

DB 358 CAGCGCTTCAGCAATCGCTTCTTAAACGAAACCACTCGAATVCTGTCACAA 417

QY 138 ProLeuValAlaAlaGlnAlaValaSerArgGlnG|uTyraAspAlaAlaValThrAlaLys 157

DB 418 CAATGATGTCTCAAGATCAACACCTGTCAGATTTAGGTGGCTGAAGCTGCTTAT 477

QY 158 ArgSerAlaGlnAlaGlyValAlaAlaGlnAlaAlaIleLysSerAlaGlyIle--- 176

DB 478 AAACAGCTCGACGCAAGTTAAAGCATTAAGTACCAAAATGAGTCTGCAAAATTAAG 537

QY 177 -----AsnLeuAsnArgSerArgIleThrAlaProIleSerGlyPhe 190

DB 538 CGTTCAACAGACCAACCAATATTGGCTATACGCTATGTCGCCCAATGATGATACG 597

QY 191 IlleGlyInserLysValSerG|uGlyThrLeuLeuAsnAlaGlyAspThrThValLeu 210

DB 598 GTTTCGCGATTGACTGCAAGAGGTCACACGCTAAACGCA----- 639

QY 211 AlaThrIleArgGlnThrAsnProMetTyraValAsnValThrGlnSerAlaSerGluVal 230

DB 640 -----AACCAAGCGCTCCAACTATGCTCAAAATTCGAAAACTCAAAAT- 684

QY 231 MetLysLeuArgArgGlnIleAlaG|uGlyLysLeuLeuAlaAlaAspGlyValIleAla 250

DB 685 ATGACGATTTAAAGCAGCGTACGAGCCGATTTATTAAGTGTG----- 729

QY 251 ValGlyIleLysPheAspAspGlyThrValTyraProG|uLysGlyArgLeuPheAla 270

DB 730 -----GAAAAAGGTCAGCAGCTCTATTTC 753

QY 271 AspProValAlaAsnG|uSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAsp 290

DB 754 ACGACCTTAGGTATGAACCAACAGCGCTATGCAACCTTACTCAATTTGAACCTGCTCA 813

QY 291 GlnAsnIle-----LeuMetProGlyLeuTyra 299

DB 814 GATTGATCTCTAGTGAATCAACAGCACACAAGTTCAACAACAGCTGACGGTTTAC 873

QY 300 ValArgValLeuMetAsp-----GlnValAlaValAsp----- 310

DB 874 TACAACGCTTATTTGATGTTTCCAAATAAGGACGCAAAATGCGTATGATGATGATGCA 933

QY 311 -----AsnAlaPheValValProGlnGlnAlaVal 320

DB 934 CAATTTATATCGTATTAATTAATTCAGCAAAATGCTTACTGTTCCATCTTCTGCGTTA 993

QY 321 Thr----- 321

DB 994 AGTAGCAAAACATTTTCTGCGCAAGAAACCAACAGGTCAATCAGCAGATAAGCAAGT 1053

QY 322 -----ArgGlyAlaLys----- 325

DB 1054 TCTACTCCAAAGTGCAGAAACGACAGCATCAAGTTAAAGCGCCGTTTGAACGCTTAAT 1113

QY 326 -----AspThrValMet 329

DB 1114 TTAACCTCGTGAACAAAACAGCTTATTAAGCAAGCAACCAACTGAGTATGTCGT 1173

QY 330 IlleValAsnAlaGlnG|uGlyMetG|uProArgGluValThrValAlaGlnGlnGlnGly 349

DB 1174 GTTTACACAGCAGATGGTATGACTACTAAACCAAAATTTGGTATTAATTAACCGT 1233

QY 350 ThrAsnTrpIleValThrSerGlyLeuLysAspGlyAspLysValAlaValAlaGlyIle 369

DB 1234 GTAATCGCAGAGTACTGCGGATTTAAACAGGTGACCAAGTTGATTAATTCGGAT--- 1290

QY 370 SerIleAlaGlyIleThrGlyAlaLysLysValThrProLysG|uTyraLysSerGly 389

DB 1291 -----AGTTCAAGAA 1299

QY 390 AsnGlnAlaAlaAlaProGlnInserGly 398

DB 1300 AACTCTGACAGCTTCTGCAACAGTGT 1326

RESULT 28

US-09-252-991A-4039

Sequence 4039, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

PRIOR FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 4039

LENGTH: 1200

TYPE: DNA

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-4039

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
1.25e-19	1200	267.00	93	64	170	36
Percent Similarity:		43.25%				
Best Local Similarity:		25.62%				
Query Match:		13.22%				

DB: 4 Gaps: 8

US-09-889-756a-2 (1-412) x US-09-252-991A-4039 (1-1200)

QY 11 AlaAlaAlaLeuAlaAlaAlaValAlaLeuValLeuSerGlyGlyGlyAsp 30

DB 112 AGTCCCGCTCATCTGCGCGCGGTATCGGATCCGCTCAACGCCACCGCTCGCGC 171

QY 31 AlaAlaGlnGlyGlnProAlaGlyArgGluAlaProAlaProValAlaGlyVal 50

DB 172 AAGAAAGACCGCGCGCTTCGCGCGC-----TACCGCGCGTAAAGGTGCGCTGCC 225

QY 51 ThrValHisProGlnThrValAlaAlaLeuThrValGluLeuProGlyArgLeuGluSerLeu 70

DB 226 TCGGTGAGGCGGCGGTGCGCGCGCTTCGATGATGCGCGCGCGAGTGGAGCGCGT 285

QY 71 ArgThrAlaAspValArgAlaGlnValGlyIleIleGlnLysArgLeuPheGlnGlu 90

DB 286 CGCGAGTCCAGGTGCGCGCGGAAAGCGGAGCGGATCCCGCATCGCTTCGAAATCG 345

QY 91 GlySerTyraValArgAlaGlyGlnProLeuTyraGlnIleAspSerSerThrTyraGluAla 110

DB 346 GCGCAGAGGTGCGCAAGGCGACTTGTGTGCTGCACTCAACAGCGCGGTGGAACAGGCC 405

QY 111 AsnLeuG|uSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaLysAlaAsp 130

```

Db 406 GAGCTGATCCGCTCAAGCCGAGTTGCGCAATGCCAGATCTCTCATGCCCGTCCG--- 462
Qy 131 AlaAspLeuAlaArgTyrIleuProLeuValAlaIleValSerArgGlnIleuTyr 150
Db 463 -----CGCAAGCTGGTAGAGCCCAACGTGCGCTCCAGGAAACAGCTG 504
Qy 151 AspAlaIleValThrAlaIleArgSerIleGluIleValIleValAlaGlnAla 170
Db 505 GACAAAGCGCTGCGCCGCGCAATGCGCTGCGCGCGCGCGCAAGCCAGCGCTG 564
Qy 171 IleuSerIleAlaIleAsnLeuAsnArgSerArgIleThrAlaProIleSerGlyPhe 190
Db 565 ATC-----GACCAAGAGCGCATCCGCGCGCTCTCTCCGCGAG 603
Qy 191 IleGlyIleSerIleValSerGluIleThrIleuLeuAsnAlaGlyAspThrThrValLeu 210
Db 604 CTCGGCATCCCGCGGTGACCTCGGCAATCCTCGCGCTGCGCGCGCGCGCGCGCG 663
Qy 211 AlaThrIleArgIleThrAsnProMetTyrValAsnValThrGlnSerIleSerGluVal 230
Db 664 CTGGTGATGCGGACCGCTGAAAGCAATTTCTCCGAGCAAGAACACAGCTCCGAG 723
Qy 231 MetIleuLeuArgArgGlnIleAlaGluIleuLeuAlaIleAspGlyValIleAla 250
Db 724 CTGAAGCTCGCCAGCCC-----CTCGAG 747
Qy 251 ValGlyIleLeuPheAspAspGlyThrValTyrProGluIleuValArgLeuPheAla 270
Db 748 GTCCGTGTCGACGGCTATCCGGGGCGAGCTTCCG-----GGCGCATGAGCGCATC 801
Qy 271 AspProValValAsnGluSerThrGlyGlnIleThrIleuArgAlaIleAlaProAsnAsp 290
Db 802 GACCGCTGATCGGACAGTCCGCG---ACGGTGCAGGTCCAGCGCTTGTGACAAACCCC 858
Qy 291 GlnAsnIleLeuMetProGlyIleuTyrValArgValleuMetAspGlnValAlaValAsp 310
Db 859 GAAGCGCTGCTCGCCCGCGCATGTTCCGACAGCTCGGCTCGCGCAAGCCGAGCGG 918
Qy 311 AsnAlaPheValValProGlnIleAlaValThrArgIleAla---LysAspThrValMet 329
Db 919 CCGTGGTGAAGCGTCCGCGAAGCCGCTATCCGCTACCGCTTACGGCGACACCTGTTC 978
Qy 330 IleValAsnAlaGlnIleGlyMetGluProArgIleValThrValAlaGlnGlnIleGly 349
Db 979 GTCCGCCACAGAGAGCGGACCGCGCTCAGCCCAAGCGCTCGGTGCGGATCGGC 1038
Qy 350 ThrAsnTrp-----IleValThrSerGlyLeuIleAspGlyAspValVal 365
Db 1039 GAGCGCTGAGCGGTGCGGTGAATCTCCAGGGGCTCGCGCGAGCGGAGTGTG 1098
Qy 366 ValGluGly 368
Db 1099 ACTTCCGGA 1107

RESULT 29
US-09-252-991A-4005
; Sequence 4005, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIORITY FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIORITY FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4005
; LENGTH: 1398
; TYPE: DNA

```

```

; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4005

Alignment Scores:
Pred. No.: 1.59e-19 Length: 1398
Score: 267.00 Matches: 93
Percent Similarity: 43.25% Conservative: 64
Best Local Similarity: 25.62% Mismatches: 170
Query Match: 13.22% Indels: 36
DB: 4 Gaps: 8

US-09-889-756a-2 (1-412) x US-09-252-991A-4005 (1-1398)

Qy 11 AlaAlaIleuAlaIleAlaValAlaIleuValLeuSerSerGlyIleGlyIleValAsp 30
Db 32 AGTCCCGCTCATCTCCGCGCGCGGTATCGCATCCCGCTTACCGCCAGCGCGCGCG 91
Qy 31 AlaAlaGlnIleGlyIleProAlaGlyIleValAlaProAlaProValIleVal 50
Db 92 AAGAAAGACCGCGCGGTTCGCGCGC-----TACCGCGGTAAAGTCCCTCGCC 145
Qy 51 ThrValHisProGlnThrValAlaIleuThrValGluIleuProGlyIleGluIleu 70
Db 146 TCGGTGAGCGCGCGGTGTCGCGCGCTTCTGATGCGCTGCGAGCTGAGCGCGGT 205
Qy 71 ArgThrIleAspValArgAlaGlnIleValGlyIleIleGlnIleuArgLeuPheGlnIle 90
Db 206 GCGCAGTCCAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 265
Qy 91 GlySerTyrValArgAlaGlyIleProIleuTyrGlnIleAspSerSerThrTyrGluAla 110
Db 266 GCGCAGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 325
Qy 111 AsnIleuSerIleAlaArgIleAlaIleuAlaIleuAlaIleuAlaIleuAlaIleu 130
Db 326 GAGCTGATCCGCTCAAGGCGCGAGTTCGCAATGCGAGATCCTCCATGCGCGTGC 382
Qy 131 AlaAspLeuAlaArgTyrIleuProLeuValAlaIleuAlaIleuAlaIleuAlaIleu 150
Db 383 -----CCCAAGCTGTGAGACCGCGCAACCTCGCGCGCGCGCGCGCGCGCG 424
Qy 151 AspAlaIleValThrAlaIleuArgSerIleAlaGlyIleValIleValAlaIleuAla 170
Db 425 GACAAAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 484
Qy 171 IleuSerIleAlaIleuAsnArgSerArgIleThrAlaProIleSerGlyPhe 190
Db 485 ATC-----GACCAAGAGCGATCCGCGCGCTTCTCTCGCGCGAG 523
Qy 191 IleGlyIleSerIleValSerGluIleThrIleuLeuAsnAlaGlyAspThrThrValLeu 210
Db 524 CTCGGCATCCCGCGGTGACCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 583
Qy 211 AlaThrIleArgIleThrAsnProMetTyrValAsnValThrGlnSerIleSerGluVal 230
Db 584 CTGGTGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 643
Qy 231 MetIleuLeuArgArgGlnIleAlaGluIleuLeuAlaIleuAlaIleuAlaIleu 250
Db 644 CTGAAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 667
Qy 251 ValGlyIleLeuPheAspAspGlyThrValTyrProGluIleuValArgLeuPheAla 270
Db 668 GTCTGGTGAAGCGCTATCCGGGGCGAGCTTCCG-----GGCGCATGAGCGCATC 721
Qy 271 AspProValValAsnGluSerThrGlyGlnIleThrIleuArgAlaIleAlaProAsnAsp 290
Db 722 GACCGCTGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 778
Qy 291 GlnAsnIleuLeuMetProGlyIleuTyrValArgValleuMetAspGlnIleAlaValAsp 310
Db 779 GAAGCGCTGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 838

```

```

QY 311 AsnAlaPheValProGlnGlnAlaValThrArgIValA---LysAspThrValMet 329
   ::::::::::::::::::::
Db 839 CCGTCCGAGACCGTCCGGAACCGCGTCACTATACCGCTACGGGAGACCGGTTC 898
QY 330 IleValAsnAlaGlnGlyGlyMetGluProArgIValThrValAlaGlnGlnGly 349
   ::::::::::::::::::::
Db 899 GTCCCCACGACGAGCGGAGCCGCGCTCAGCCGCAAGCGGTCTCGGTCCGATCCGC 958
QY 350 ThrAsnTrp-----IleValThrSerGlyLeuLysAspGlyAspIleValVal 365
   959 GAGCGCTGGAGCGGTCCGCTGGAATCTCCAGGCGCTCGCCGAGGCGGAGGTAGTG 1018
Db
QY 366 ValGlnGly 368
   1019 ACTCCGGA 1027
Db

RESULT 30
US-09-252-991A-3906/c
; Sequence 3906, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 3906
; LENGTH: 4704
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-3906

Alignment Scores:
Pred. No.: 1,08e-18 Length: 4704
Score: 267.00 Matches: 93
Percent Similarity: 43.25% Conservative: 64
Best Local Similarity: 25.62% Mismatches: 170
Query Match: 13.22% Indels: 36
DB: 4 Gaps: 8

US-09-889-756a-2 (1-412) x US-09-252-991A-3906 (1-4704)
QY 11 AlaAlaIleuAlaAlaAlaValAlaIleuValIleuSerSerCysGlyGlyValAsp 30
   ::::::::::::::::::::
Db 4500 AGTCCCGGCTCATCTGCGCGCGGCTTATCGGCAATCCGCTACGCCACCGGCTCGGCG 4441
QY 31 AlaAlaGlnGlyGlnProAlaGlyArgIValAProAlaProValAlaGlyValVal 50
   ::::::::::::::::::::
Db 4440 AAGAAAGCGCGCGGCTTTCGCGGC-----TACCGCGGTAAGGTGCGCTCGCC 4387
QY 51 ThrValHisProGlnThrValAlaIleuThrValGluLeuProGlyArgLeuGlnSerIeu 70
   ::::::::::::::::::::
Db 4386 TCGGTGAGCGCGCGGTGTCGCGCGCTCTTCGATGCGGTGCGGAGCTGAGGCGGT 4327
QY 71 ArgThrAlaAspValArgAlaGlnValGlyIleIleGlnLysArgLeuPheGlnGlu 90
   ::::::::::::::::::::
Db 4326 CGCCAGTCCAGGTGCGCGCGGAGCGGAGCGGATCACCAGATCGCTTCGAATCG 4267
QY 91 GlySerTyrValArgAlaGlyGlnProLeuTyrGlnIleAspSerSerThrTyrGlnAla 110
   ::::::::::::::::::::
Db 4266 GCGCAGAGGTGCGGAGCGGCAAGTCTGTCGCACTCAACGACGCGGTGGAACAGCGC 4207
QY 111 AsnLeuGlnSerAlaArgAlaGlnIleuAlaThrAlaGlnAlaThrLeuAlaIleAsp 130
   ::::::::::::::::::::
Db 4206 GAGCTGATCCCTCTCAAGCGCGAGTGGCAATGCGGATCTCCATGCGCGCG--- 4150
QY 131 AlaAspLeuAlaArgTyrLysProLeuValAlaIleuAlaValSerArgGlnGlnIuTyr 150

```

```

Db 4149 -----CCGAACTGTAGACGCAACGTCGCTCGCAGAACACTG 4108
   ::::::::::::::::::::
QY 151 AspAlaIleValThrAlaLysArgSerAlaGlnAlaGlyValLysAlaIleAlaAla 170
   ::::::::::::::::::::
Db 4107 GACAACTGCTGCGCGCGGAGCATGCGCTCGCGGTGCGGACAGACCCAGCGCTG 4048
QY 171 IleYsSerAlaGlyIleAsnLeuAsnArgSerArgIleThrAlaProIleSerGlyPhe 190
   4047 ATC-----GACCAAGGCGATCCGCGGCTTCTCGGCGAG 4009
Db
QY 191 IleGlyGlnSerLysValSerGlyGlyThrLeuLeuAsnAlaGlyAspThrThrValLeu 210
   4008 CTCGCGATCCGCGCGCGGACCTTCGCGAGTACTTCGCGCGCGCGCGCGCGCGAGC 3949
Db
QY 211 AlaThrIleArgGlnThrAsnProMetTyrValAsnValThrGlnSerAlaSerGluVal 230
   3948 CTGCTGATGCGCGGACCTGAAAGCAATTTCTCTCGAGCGAAGACCAAGTCCGAG 3889
QY 231 MetLysLeuArgArgGlnIleAlaGlnGlyLysLeuLeuAlaIleAspGlyValIleAla 250
   3888 CTGAAGCTCGGCGGCGCG-----CTCGAG 3865
Db
QY 251 ValGlyIleLysPheAspAspGlyThrValTyrProGluLysGlyArgLeuLeuPheAla 270
   3864 GTCTGTGTCGACGCGCTATCCGCGGCGGCGAGCTTCCG-----GCGCGCATCAGCGGCATC 3811
Db
QY 271 AspProValValAsnGlnSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAsp 290
   3810 GACCGCTGATCGGCAAGTGGCGC-----ACGTCGAGGTCCAGGCGCTTTCGAGACACCCC 3754
Db
QY 291 GlnAsnIleLeuMetProGlyLeuTyrValArgValLeuMetAspGlnValAlaValAsp 310
   3753 GAAAGCTGCTCGCGCGCGCGGATGTCCAGCATCCGGGTCTCGCCCAAGCGGAGCGG 3694
QY 311 AsnAlaPheValValProGlnGlnAlaValThrArgIValA---LysAspThrValMet 329
   ::::::::::::::::::::
Db 3693 CCGTCCGAGAGCGGCGCGGAAACCGCGGTCACTATACCGCTACGCGGAGACCGGTTC 3634
QY 330 IleValAsnAlaGlnGlyGlyMetGluProArgIValThrValAlaGlnGlnGly 349
   3633 GTCCCCACGACGAGCGGAGCGCGCTCAGCCGCAAGCGGTCTCGGTGCGATCGGC 3574
Db
QY 350 ThrAsnTrp-----IleValThrSerGlyLeuLysAspGlyAspIleValVal 365
   3573 GAGCGCTGGAGCGTCCGCTGGAATCTCCAGGCGCTCCCGGAGGCGAGCGGTAAGT 3514
Db
QY 366 ValGlnGly 368
   3513 ACTCCGGA 3505
Db

RESULT 31
US-09-252-991A-12042/c
; Sequence 12042, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12042
; LENGTH: 1527
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12042

```


Alignment Scores:

Pred. No.: 1,55e-18 Length: 1527
 Score: 258.50 Matches: 107
 Percent Similarity: 43.99% Conservative: 76
 Best Local Similarity: 25.72% Mismatches: 171
 Query Match: 12.80% Indels: 62
 DB: 4 Gaps: 13

US-09-889-756A-2 (1-412) x US-09-252-991A-12042 (1-1527)

QY 10 ArgAlaAlaLeuAlaAlaAlaValAlaLeuValLeuSerSerCysGlyValGly 29
 Db 1485 CGAGTGGGCTGCGCCGATCTCTGCGCGGCTGCTGCGTGGG----- 1435
 QY 30 AspaAlaAglnglyGlnProAlaGlyArgGluAlaProAlaProValAlaGlyVal 49
 Db 1434 -----GCCAGCGCGCGCGCGAGAACAC 1411
 QY 50 ValThrValHisProGlnThrVal-----AlaLeuThrValGluLeuPro 64
 Db 1410 GTCCGTGTGCTGGCGACAGCGTGAAGATGGCCGAGTTCCTGCGCCACCTGCATCACC 1351
 QY 65 GlyArgLeuGluSerLeuArgThrAlaAspValArgAlaGlnValGlyGlyIleIleGln 84
 Db 1350 GGGCAATCCAGCGACGGTACAGCGCCGACGTCGTCTGCTGCGCGCAAGATCTGC 1291
 QY 85 LysArgLeuPheGlnGlnGlySerTyValArgAlaGlyGlnProLeuTyArgIleAsp 104
 Db 1290 GAGGCGCTGTCATGTCGCGACACCACTGCGCTGCGTCCAGGTCGTGCGCGCTGCAC 1231
 QY 105 SerSerThrTyGluAlaAsnLeuGluSerAlaArgAlaGlnLeuAlaThrAlaGlnAla 124
 Db 1230 CCGCAGACGACGCGCAACAGTGAAGACGCCGCGCGCGCGCGCAGCAGCGC 1171
 QY 125 ThrLeuAlaLysAlaAspAlaAspLeuAlaArgTyIleProLeuValAlaAlaGluAla 144
 Db 1170 CAGTCGAGCTGCGCGACCTCACTACACCGACGAGAGCGCTGCTGCCAAGGCTTAC 1111
 QY 145 ValSerArgGlnGluTyArgAspAlaAlaValThrAlaLysArgSerAlaGluAlaGlyVal 164
 Db 1110 ACCAGCCAGACCGATGCGACAGCGCGCTGCTGCTGCGCGCGCGCAGAGTTCCTG 1051
 QY 165 LysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsnArgSerArgIleThr 184
 Db 1050 AAGGCGCGCCAGGCGAGTGGCCACGCGCGACCTGCTTCCTATACGAGCTGCTG 991
 QY 185 AlaProIleSerGlyPheIleGlyGlnSerLysValSerGluGlyThrLeuLeuAsnAla 204
 Db 990 GCTCCGACCGCGGGTCACTGCGCGCGCGACGAGTCCGCGCAGGTGCGCAGGCGC 931
 QY 205 GlyAspThrThrValLeuAlaThrIleArgGln-----ThrAsnProMetTyValAsn 222
 Db 930 -----ACGTCGCCGATCTTCACTCCCGCGCGACGCGCGACGCGCGTTCAC 877
 QY 223 ValThrGlnSer-----AlaSerGluValMetLysLeuArgArgGlnIleAlaGly 240
 Db 876 GTCAAGAGTGTGTTGTTACGACCAAGTGTGACGCGCGACGAGTACCGCTCAC 823
 QY 241 LysLeuLeuAlaAlaAspGlyValIleAlaValGlyIleLysPheAspAspGlyThrVal 260
 Db 822 ---CTGCTCGCAGCGCGAAGTCAACCGCC----- 796
 QY 261 TyrProGluLysGlyArgLeuLeuPheAlaAspProValAlaAsnGluSerThrGln 280
 Db 795 -----AGCGCAAGGTCCCGCGAGATACCCCGAGGTGAGCAAGCGCGGATACG 745
 QY 281 IleThrLeuArgAla-----AlaValProAsnAspGlnAsnIleLeuMetProGly 297
 Db 744 CTGAAGGTCAAGGTCCGCTCGACGCGTCCGCGGAAATGAGCTC-----GGC 694
 QY 298 LeuTyValArgValLeuMetAspGlnValAlaValAspAsnAlaPheValValProGln 317
 Db 693 AGCGGTGTCAACGCCAGCGTCCGCGCGCGCC---GAGCACAGCGTGTGCTGCTGCG 637

QY 318 GlnAlaValThrArg---GlyAlaLysAspThrValMetIleValAsnAlaGlnGly 336
 Db 636 TCGCGCTGTCCAGAGTCCGCGACGACCGCGCGCTGCTGCTGCACCGACGAGGAC 577
 QY 337 MetGluProArgGluValThrValAlaGlnGlnGlnGlyThrAsnThrIleValThrSer 356
 Db 576 GCGCGTGTCAACCGGTGCGCGGTGGCAGCTACCGCAGAGAAAGGTGTCATGACGT 517
 QY 357 GlyLeuLysAspIleAspLysValValAlaGluGly----- 368
 Db 516 GCGCTGAGGCGGCGCAGACGAGTGTGACGCGTGGCGCCCACTGCTCATCGCGGAC 457
 QY 369 ---IleSerIleAla-GlyIleThrGlyAlaLysValThrPro-----LysGluTr 385
 Db 456 GTGCTGAGGTGCGCCAGCGCGCGCGACCGCAGCCACGACCGCCAGCGCGCGT 397
 QY 385 PalAspSerGluAsnGlnAlaAlaAlaProGlnSerGlyValGln 400
 Db 396 GCGCGAGCGCAGCATGAAGCGTTCCTTCCCTCGCGCGCTGTTGCG 351

RESULT 32

US-09-252-991A-11939
 ; Sequence 11939, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; PRIOR FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 11939
 ; LENGTH: 2313
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 ; US-09-252-991A-11939

Alignment Scores:

Pred. No.: 2.99e-18 Length: 2313
 Score: 258.50 Matches: 107
 Percent Similarity: 43.99% Conservative: 76
 Best Local Similarity: 25.72% Mismatches: 171
 Query Match: 12.80% Indels: 62
 DB: 4 Gaps: 13

US-09-889-756A-2 (1-412) x US-09-252-991A-11939 (1-2313)

QY 10 ArgAlaAlaAlaLeuAlaAlaAlaValAlaLeuValLeuSerSerCysGlyValGly 29
 Db 140 CGAGTGGGCTGCGCCGATCTCTGCGCGGCTGCTGCGTGGG----- 190
 QY 30 AspaAlaAglnglyGlnProAlaGlyArgGluAlaProAlaProValAlaGlyVal 49
 Db 191 -----GCCAGCGCGCGCGCGAGAACAC 214
 QY 50 ValThrValHisProGlnThrVal-----AlaLeuThrValGluLeuPro 64
 Db 215 GTCCGTGTGCTGCGCGACGAGTGAAGATGGCCGAGTTCGCTGCGCACCTGCATCACC 274
 QY 65 GlyArgLeuGluSerLeuArgThrAlaAspValArgAlaGlnValGlyGlyIleIleGln 84
 Db 275 GCGCAATCCAGCGACGCGGTACAGGCCGACGATGTCCTGCTGCGCGCAAGATCTGC 334
 QY 85 LysArgLeuPheGlnGlnGlySerTyValArgAlaGlyGlnProLeuTyArgIleAsp 104
 Db 335 GAGCGCTGTGCTGATGCGCGACGATGCGCGCTGCGCAGAGTCTGCGCGCTGCGAC 394


```

Db 778 ---CTGCTCGGCAAGCCGGAAGTCAACGGC----- 804
Qy 261 TyrProGluYsglyArgLeuLeuPheAlaaspProValValAsnGluSerThrGlyGln 280
Db 805 -----AGCGGCAAGAGTCCGCGAGATCAACCCGAGGTGAGACGCGACGCGGTACG 855
Qy 281 IleThrLeuAlaGla-----AlaValProAsnAspGlnAsnIleLeuMetProGly 297
Db 856 CTGAAGGTCAGAGTGGCGCTCGACTGCGTGGCGGCGGAAATGAGCTC-----GGC 906
Qy 298 LeuTyrValArgValLeuMetAspGlnValAlaValAspAsnAlaPheValValProGln 317
Db 907 AGCGTGTCAACGCGAGCGTCCGCGCGCGCGC-----GAGCAAGCGGTGTCTGCTCG 963
Qy 318 GlnAlaValThrArg---GlyAlaLysaspThrValMetIleValAsnAlaGlnGlyGly 336
Db 964 TCGGCGGTCTCAAGGTGCGGAGCGCGCGGTCTGTCTGCTGACACGACGAGCGAAG 1023
Qy 337 MetGluProArgGluValThrValAlaGlnGlnGlnGlyThrAsnTyrIleValThrSer 356
Db 1024 GCGCGTCTGCAACGCGTGGCGGTGCGACGCTACGCGAGGAGAGGTGTCATCGACGCT 1083
Qy 357 GlyLeuLysaspGlyAspLysValValValGluGly 368
Db 1084 GGCGTGAAGCGCGGCGACAGCGGTGTCACGGTGGCG 1119

RESULT 34
US-09-252-991A-7845
; Sequence 7845, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 7845
; LENGTH: 1011
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7845

Alignment Scores:
Pred. No.: 2,85e-18 Length: 1011
Score: 253.50 Matches: 90
Percent Similarity: 46.56% Conservative: 59
Best Local Similarity: 28.12% Mismatches: 120
Query Match: 12.56% Indels: 51
DB: 4 Gaps: 12

US-09-889-756a-2 (1-412) x US-09-252-991A-7845 (1-1011)
Qy 14 LeuAlaAlaAlaValAlaLeuValLeuSerSerCysGly---LysGlyGlyAspAlaAla 32
Db 95 CTCGCCCGCGGTGATCGCGGTGTCGATTCCTCGCCGCGTACAAAGGCTCTCATCCATCCGT 154
Qy 33 GlnGlyGlnProAlaGlyArgGluAlaProAlaProValAlaGlyValValThrVal 52
Db 155 CAG-----CAGATCGCCCTTTTCAGCGCACCGAAACCGCGATCAGCGTG----- 199
Qy 53 HisProGlnThrValAlaLeuThrValGluLeuPro----- 64
Db 200 -----ACCGCAGCCTGCGCGGAAAGCGTCTCGGAGAGCGCGCTGCCAGCCATC 250
Qy 65 GlyArgLeuGluSerLeuArgThrAlaAspValAlaArgAlaGlnValGlyGlyIleLeuGln 84
Db 251 GGAGGCTCAAGGATTCAGAGGGCGTGACCTCACCGCGAAGTCTCCGACGCGTACG 310

```

```

Qy 85 LysArgLeuPheGlnGlnGlySerTyrValArgAlaGlnProLeuTyrGlnIleAsp 104
Db 311 GAGTACTGTTCTTCCCGGAGCAGAGTGAAGTGAACCAACCGCTGATCATCGTTGAA 370
Qy 105 SerSerThrTyrGluAlaAsnLeuGluSerAlaArgAlaGlnLeuAlaThrAlaGlnAla 124
Db 371 AGCAGCTCGAGAGGACCATCTCGC-----ACTGCCAGAGCC 409
Qy 125 ThrLeuAlaLysAlaAspAlaAspLeuAlaArgTyrLysProLeuValAlaAlaGlnAla 144
Db 410 GATTCGCGCTGCGCAGGCGCGAGTACAGCGCGCGCGAAGTGAATCGAGCAGAGGCC 469
Qy 145 ValSerArgGlnGlnTyrAsp-----AlaAlaValThrAlaLysAspSerAlaGlnAla 162
Db 470 ATCTGAAAAGCGAATTCGATCTGTCGCGCGCGAGGCGCGAAGACGAGCCACCGTC 529
Qy 163 GlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsnArgSerArg 182
Db 530 -----GCCAGCTGAAGCGCGC-----CTGGCGAAGAGCGC 562
Qy 183 IleThrAlaProIleSerGlyPheIleGlyGlnSerLysValSerGlyThrLeuLeu 202
Db 563 GTGCTCGCGCTTTCGCGGAGCATCGCATCGCGAGGTGACGTCGCGGACTACGATC 622
Qy 203 AsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAsnProMetTyrValAsn 222
Db 623 TCGCGCGCG-----ACCGCATCGCATCTTGAGAGACCTTTCACCTGCTCTCGAT 676
Qy 223 ValThrGlnSerAlaSerGluValMetLysLeuArgArg---GlnIleAlaGlnGlyLys 241
Db 677 TTCACCTCGCGCGAGCAGACTTCCCTGCTCAGCGCGCGAGCTGTGTAAGTCCGG 736
Qy 242 LeuLeuAlaAlaAspGlyValIleAlaValGlyIleLysPheAspArgGlyThrValTyr 261
Db 737 GTCGCGCTTACCCCGCGCAGGTG-----TTCAGC----- 766
Qy 262 ProGluLysGlyArgLeuLeuPheAlaaspProValValAsnGluSerThrGlyGlnIle 281
Db 767 -----GCCAGATCGCGCGCGCATCAACCCAGGTGACCAACGAGACCGGACCTG 817
Qy 282 ThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetProGlyLeuTyrValArg 301
Db 818 CAGTCCGCGCTGCTCGGAGAACCGGACGGAAGCTGCTGCGGCGATGTCGCCAAC 877
Qy 302 ValLeuMetAspGlnValAlaValAspAsnAlaPheValValProGlnGlnAlaValThr 321
Db 878 CTCGAGGTGATGTTGCTGCGCGAGGAAACAACGCGTGTGTCGCGGAGCGGATCACCC 937

RESULT 35
US-09-252-991A-4631
; Sequence 4631, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4631
; LENGTH: 1170
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4631

Alignment Scores:
Pred. No.: 2.68e-17 Length: 1170

```


Db 296 CTCGATCCCGACGAGCGCTGCTGCACTGAGGCGGCGCGCCAGCTGATGCGCCG 355
Qy 123 GlnAlaThrLeuAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 142
Db 356 GAGGCGCACTTGTGACAGCGGTGCGCGAGTACCGGCTTACCGCACTTGTGACCGC 415
Qy 143 GlnAlaValSerArgGlnGluTyrAspAlaAlaAlaAlaAlaAlaAlaAlaAla 162
Db 416 AACCTGGTCACTTCCAGTTCGAGAACATCCAGAACACTACCGCGCGGAGGCGC 475
Qy 163 GlyValLysAlaAlaGlnAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 182
Db 476 CGGTGAGGAGCATCTCCGCGCAATTCACCTCCCGCAACAGCGCGGTACGCGTG 535
Qy 183 IleThrAlaProIleSerGlyPheIleGlyGlnSerLysValSerGluGlyThrLeu 202
Db 536 CTGGCTGCGCCCGAGATGCGGTATCGCCAGCGCGCGCTGAGGTGGCCAGGTGG 595
Qy 203 AsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAsnProMetTyrVal 222
Db 596 GCGGCGCGAGACGAGCTTTCAGCTGCGCGCGAGCGCAACGCGAGGTCTGATCGGC 655
Qy 223 ValThrGlnSerAlaSerGluValMetLysLeuArgGlnIleAlaGluGlyLys 242
Db 656 CTGCGGAGACACACTTCGCAAGCTTCGCACTGCGCGAGCGCGGTGCGTGAATCTGG 715
Qy 243 LeuAlaAlaAspGlyValIleAlaValGlyIleLysPheAspAspGlyThrValTyr 262
Db 716 TCGCAAGCGCGACAGCTTTCGCGCGGATTCGCGAGCTCTGCGCGCGCGGATCG 775
Qy 263 GluLys-----GlyArgLeuLeuPheAlaAspProValValAsnGluSerThr 278
Db 776 CAATCGGTACTCTGCGCGCGGAGTTCGCGAGCTTCGCGAGCTCTGCGCGCGGATCG 835
Qy 279 GlyLys-----IleThrLeuArgAlaAlaAlaAlaAlaAlaAlaAlaAla 293
Db 836 GCGCGAGCGCGCGGTCTGAGTGTGCGCGCGCGCGAGCGCGGTGCGCG----- 880
Qy 294 LeuMetProGlyLeuTyrValArgValLeuMetAspGlnValAlaValAspAlaPhe 313
Db 881 -----TTA 883
Qy 314 ValValProGlnGlnAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 333
Db 884 TCGGTTCCTTGTGCGCGGTACCGCGAGCGCGCGCGGTCTGCGGTG----- 937
Qy 334 GlnGlyMetGluProArgGlnValThrValAlaGlnGln----- 347
Db 938 -----GTGAGCGCGCGCGAGCTGCGCGCGCGCGGTGCGCGCGGTGCG 988
Qy 348 ---GlnGlyThrAsnTrpIleValThrSerGlyLeuLysAspGlyAspLysValVal 366
Db 989 TATCCGAGAGCGCGGTGCGCGGTCTGAGAGGCTGAGGCTGCGGTGCGGTGCGC 1048
Qy 367 GlnGlyIleSerIleAlaGlyIleThrGlyAlaLysValThrProLysGluTyrAla 386
Db 1049 ACGGCGGTGCGAGT-----CTTGGGAGGAGGAGGAGGTGCGGTGCGAGCGGCGC 1102
Qy 387 SerSerGluSerGlnAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 396
Db 1103 AACCGACGCGTGAACCTGCGCGCGCAAGAG 1132
RESULT 37
US-09-252-991A-14157/c
; Sequence 14157, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 14157
; LENGTH: 588
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14157
Alignment Scores:
Pred. No.: 1,16e-17 Length: 588
Score: 244.50 Matches: 56
Percent Similarity: 56.83% Conservative: 23
Best Local Similarity: 40.29% Mismatches: 55
Query Match: 12.11% Indels: 5
Gaps: 3
US-09-889-756a-2 (1-412) x US-09-252-991A-14157 (1-588)
Qy 261 TyrProGluLysGlyValLeuLeuPheAlaAspProValValAsnGluSerThrGly 280
Db 587 TACGAGCGCGAGCGCGGTGCGAGTTCGCGAGGTGCGGTGCGGTGCGGTGCGGTG 528
Qy 281 IleThrLeuArgAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 300
Db 527 ATGCGCTGCGCGCGCGAGTTCGCGAGTTCGCGAGTTCGCGAGTTCGCGAGTTCG 468
Qy 301 ArgValLeuMetAspGlnValAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 320
Db 467 CCGGTACGTGCGCGCGCGAGTTCGCGAGTTCGCGAGTTCGCGAGTTCGCGAGTTCG 408
Qy 321 ThrArgGlyAlaLysAspThr-----ValMetIleValAsnAlaGlnGlyMetGlu 338
Db 407 CACCGCTTCGCGCGCGAGCGCGCGCGCGAGTGTGCGCGCGCGCGCGCGCGCG 348
Qy 339 ProArgGluValThrValAlaGlnGlnGlnGlyThrAsnTrpIleValThrSerGly 358
Db 347 TCGCGAGCGTCCGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 288
Qy 359 LysAspGlyAspLysValValAlaAlaGluGlyIleSerIleAlaGlyIleThrGly 378
Db 287 GAGCGCGGTGACCGGCGGTATGCGCGCG-----CTGCTGCGGTGCGAGCGCGGTG 234
Qy 379 LysValThrProLys---GluTrpAlaSerSerGluSerGlnAlaAlaAlaAlaAla 396
Db 233 AAGATCGTCCGAGCGCGAGTGTGCGCGAGCGCGAGCGCGAGCGCGCGCGCA 177
RESULT 38
US-09-252-991A-11448/c
; Sequence 11448, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11448
; LENGTH: 465
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11448
Alignment Scores:
Pred. No.: 3.19e-17 Length: 465


```

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9189
; LENGTH: 1353
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-9189

Alignment Scores:
Pred. No.: 1,41e-14 Length: 1353
Score: 221.50 Matches: 101
Percent Similarity: 40.58% Conservative: 67
Best Local Similarity: 24.40% Mismatches: 184
Query Match: 10.97% Indels: 62
Gaps: 12

US-09-889-756A-2 (1-412) x US-09-252-991A-9189 (1-1353)

Qy 35 G1yGlnProAlaGlyArgGluAlaProAlaProValAl-GlyValValThrValHisPr 54
Db 1198 GGGCGCGTGGAGCGCCATCCGTTCCCGCAAGACAGCTTGATCCGTCAGCGCCG 1139

Qy 54 oGlnThrValAlaLeuThrValGluLeuProGlyArgLeuGluSerLeuArgThrAlaAs 74
Db 1138 CGGAGATCGAGAGCAGCGCTCGGCGCTCGGACCTCGAACCGGCGGCTACGTGA 1079

Qy 74 pValArgAlaGlnValGlyLeuLeuGlnValArgLeuPheGlnGluGlySerTyra 94
Db 1078 CGTGGCGCGCGCGCTCGGCGAGATCCGAACTTCACGTCCAGCGCGCGGAGATGT 1019

Qy 94 lAspAlaGlyGlnProLeuTyrglnLeuAspSerSerThrTyrglnAlaAsnLeuGluSe 114
Db 1018 GAGCGAAGGCGATGTGTGTGTCAGATCGACCCCTCCACCCAGCAGCGCCAAAGTCATGC 959

Qy 114 rAlaArg-----AlaGlnLeuAlaThrAlaGlnAlaThrLeuAl 127
Db 958 CGCGCGCTATTCATCGATCGATGCTCAAGCGCCAGCTGCGGAGCAAGTCGCAATCAC 899

Qy 127 alysAlaAspAlaAspLeuAlaArgTyryspProLeuValAlaAlaGluAlaValSerAr 147
Db 898 CTTGGCGCGCGCAGATGACGAGCGCGCGCTGGCGCGCGCGCGCGCGCGCTAC 839

Qy 147 gGlnGluTyraAspAlaAla-----ValThrAlaAlaArg-----SerAl 160
Db 838 CGAGAGAGTGCAGAGCGCCAGCGCGCGAGATGCTGCCACAGGACGATCGAGATGTA 779

Qy 160 agluAlaGlyValIlyAlaAlaGlnAlaAlaIlelySerAlaGlyIleAsnLeuAsnAr 180
Db 778 CAGGCGCGCAGATCGCGAGCGCGCGCTGTCGCGAGCAGCAGCGCGCGCTGCGCTA 719

Qy 180 gSerArgIleThrAlaProIleSerGlyPheIleGlyGlnSerIyValSerGluGlyTh 200
Db 718 TACCGCGATTAAGCGCGCGATGTCGCGCGAGTGTGGCGGTCGATGCGCGAGGCGCA 659

Qy 200 rLeuLeuAsnAla---GlyAspThrThrValLeuAlaThrIleArgGlnThrAsnProMe 219
Db 658 GACCTCAATGCCAGCAGCAGACAGACCCCGTTGATCTCGGATTCGCCAATTTGCGCGAT 599

Qy 219 tTyraAsnValThrGlnSerAlaSerGluValMetIyLeuArgArgGlnIle--Al 238
Db 598 GACGCTGTGGCCAGCGCTTCGGAAGCGCAGATCGCGGATCAAGCCGCGCATGCGCGC 539

Qy 238 agluGlyIyLeuLeuAlaAlaAspGly-----ValIle 249
Db 538 CTACTTCACACCTTCAGCGCGAGAGCGCGCGCTGAGCCGCGCAAGGTCCGCGAGATCCT 479

```

```

Qy 249 eaIaValGlyIleIys-----PheAspAspGlyThrValTyrrProGluIySgl 265
Db 478 CCGGTGCGCGCCAGAGCGCGTGCAGCAGCAGCAGCAGCGCGCGCGCGCGCGCGCG 419

Qy 265 yArgLeuLeuPheAlaAspProValValaAsnGluSerThrGlyGln----- 280
Db 418 C-----AGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 392

Qy 281 -----IleThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetProGl 297
Db 391 GGTGTGCTGTATACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 332

Qy 297 yLeuTyraIyValArgValLeuMetAspGlnValAlaValaAspAsnAlaPheValValProGl 317
Db 331 AATGACCGCGCGAGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 272

Qy 317 nGlnAlaValThrArgIyAlaIlys-----AspThrValMetIleValaAsnAlaGlnGl 335
Db 271 CGCGCGCATTCACGAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 212

Qy 335 yGlyMetGluProArgGluValThrValAlaGlnGlnGlnGlyThrAsnTrpIleValTh 355
Db 211 CAAGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 152

Qy 355 rSerGlyLeuIyAspGlyAspIyValValValaValaGlnGlyIleSerIleAlaGlyIleTh 375
Db 151 TGCCTGCTGAGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 118

Qy 375 rGlyAlaIyValValThrProIyGluTrpAlaSerSerGluAsnGlnAlaAlaApr 395
Db 117 -GCCGCTCCGAGCGAGCAGCAGCTGATGGAAGAACCGCGCGCGCGCGCGCGCGCG 65

Qy 395 oGlnSerGlyValGlnThrAlaSerGluAlaIySthrAla 408
Db 64 CTGATCAACTGCGCGACATTCGCAAGCGCTACCGCGCGCA 25

```

Search completed: September 8, 2003, 07:58:18
Job time : 2330 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 8, 2003, 04:57:54 ; Search time 368 Seconds
(without alignments)
3022.197 Million cell updates/sec

Title: US-09-889-756A-2
Perfect score: 2019
Sequence: 1 MAFYAFKAMRAALAAVAL.....AAPOSGVQTASEAKTASEAE 412

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Xgapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 252756 seqs, 1349719017 residues
Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xjh
-Q=Cgnt2.1/USFTO.spool/US09889756/runat.05092003.124323.23271/app.query.fasta.1.583
-DB=N.geneseq.19Jun03 -OPMT=fastlap -SUFFIX=ring -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNIT5-bits -START=1 -END=-1 -MATRIX=blonsum62 -TRANS=human40.cdi
-LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40
-MODE=LOCAL -OUTFMT=ptc -NOM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09889756 @CGN 1.1 312 @runat.05092003.124323.23271 -NCP=6 -ICPU=3
-NO_MMAP -JARSEOUTERY -NEG_SCORES=0 -WAIT -DSPBLDLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
N.Geneseq.19Jun03.*
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
25: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2019	100.0	1239	21	AAA37645	N. meningitidis BA
2	1989	98.5	1239	24	ABK54080	DNA encoding Neiss
3	1989	98.5	20352	21	AA81488	N. meningitidis pa
4	1989	98.5	349980	21	AA21611	Neisseria meningit
5	1989	98.5	837096	21	AA81489	N. meningitidis pa
6	1943	96.2	1236	25	AB239062	N. gonorrhoeae nuc
7	762.5	37.8	4879	20	AA211064	E. coli acra and a
8	701.5	34.7	7888	24	AB578861	E. coli CFT073 gen
9	631.5	31.3	99629	22	AA28550	Genomic fragment #
10	470	23.3	2967	23	AA89861	DNA encoding novel
11	459.5	22.8	2934	23	AA87973	DNA encoding novel
12	435.5	21.6	1395	21	AA65929	E. coli proliferat
13	435.5	21.6	1395	22	AA84642	E. coli growth and
14	423	21.0	1116	24	AB090066	M. capsulatus gene
15	370.5	18.4	1008	20	AA98820	Nucleotide sequenc
16	339	16.8	2112	23	AA579819	DNA encoding novel
17	339	16.8	2112	23	AA81885	DNA encoding novel
18	324	16.0	2310	23	AA88249	DNA encoding novel
19	313	15.5	1176	25	AB240695	N. gonorrhoeae nuc
20	310	15.4	1179	20	AA212306	Neisseria gonorrh
21	309.5	15.3	349980	21	AA216107	Neisseria meningit
22	309.5	15.3	143768	21	AA81490	N. meningitidis B
23	308	15.3	1176	24	AA597298	Neisseria meningit
24	308	15.3	1177	20	AA597248	Neisseria meningit
25	308	15.3	1179	20	AA212305	Neisseria meningit
26	308	15.3	1179	21	AA253631	Neisseria gonorrh
27	308	15.3	1179	21	AA253633	Neisseria meningit
28	304.5	15.1	3732	23	AA574537	DNA encoding novel
29	304.5	15.1	3732	23	AA592892	DNA encoding novel
30	304.5	15.1	3732	23	AA592892	DNA encoding novel
31	304.5	15.1	44668	21	AA81495	N. meningitidis pa
32	304.5	15.1	26923	22	AA28554	Genomic fragment #
33	292	14.5	607	24	AB24964	Oligonucleotide fo
34	292	14.5	607	24	AB24965	Oligonucleotide fo
35	286	14.2	1830121	17	AAT42063	Hemophilus influe
36	279.5	13.8	1086	20	AB090313	M. capsulatus gene
37	271	13.4	1005	20	AA212304	Neisseria meningit
38	271	13.4	1005	21	AA253632	Neisseria meningit
39	239.5	11.9	900	24	AB091605	M. capsulatus gene
40	220	10.9	607	24	AB024962	Oligonucleotide fo
41	220	10.9	607	24	AB024963	Oligonucleotide fo
42	219	10.8	569	24	AB037264	Oligonucleotide fo
43	219	10.8	569	24	AB037265	Oligonucleotide fo
44	217	10.7	306	24	AB075607	Human glycoprotein
45	188	9.3	1182	25	AB237910	N. gonorrhoeae nuc

ALIGNMENTS

RESULT 1
AAA37645
ID AAA37645 standard; DNA; 1239 BP.
XX
XX AAA37645;
AC
XX
XX 24-OCT-2000 (first entry)
DT
XX
XX N. meningitidis BASB05 coding sequence.
DE
XX
XX BASB05; diagnosis; microbial infection; invasive bacterial disease;
KW Neisseria meningitidis infection; upper respiratory tract infection;
KW bacteremia; meningitis; therapy; ss.
XX
XX Neisseria meningitidis.
OS
XX
FH Key Location/Qualifiers

FT	CDS		1..1239	
TT	/tag=	a	BASB055	
ET	/product=	BASB055		
PX				
PV	WO200043517-A1.			
PD	27-JUL-2000.			
XX				
PP	19-JAN-2000; 200OMO-EPO0425.			
PR	22-JAN-1999; 99GB-0001462.			
PR	29-JAN-1999; 99GB-0002069.			
XX	(SMIK) SMITHKLINE BEECHAM BIOLOGICALS.			
PA				
XI	Thomnard J;			
P1				
DR	WPI; 2000-476199/41.			
DR	P-PADB; AAAY0283.			
XX				
PT	Isolated BASB05 polypeptides, polymucleotides, and antibodies, the polypeptides and polymucleotides are useful as vaccines for treating and diagnosing a microbial infection such as a Neisseria meningitidis infection _			
XX				
PS	Claim 7, Page 63; 82pp; English.			
XX				
CC	This sequence encodes the Neisseria meningitidis BASB055 polypeptide of the invention. The BASB055 polypeptides and polymucleotides are useful for diagnosing and treating microbial infections such as a Neisseria meningitidis infection. They can also be used to treat any disease caused by or related to infection by a bacteria, including upper respiratory tract infection, invasive bacterial diseases (such as bacteraemia) and meningitis.			
SQ	Sequence 1239 BP; 275 A; 332 C; 384 G; 248 T; 0 other;			
Alignment Scores:				
Pred. No.:	1,48e-157	Length:	1239	
Score:	2019.00	Matches:	412	
Percent Similarity:	100.00%	Conservative:	0	
Best Local Similarity:	100.00%	Mismatches:	0	
Query Match:	100.00%	Indels:	0	
	21	Gaps:	0	
US-09-889-756A-2 (1-412)	x AAA37645 (1-1239)			
OY	1 MetAlAPheTYrAlAPheLYsaLaMeLArGLALALAleuaLAalaVaLaLeu 20			
Dd	1 ATGGCTTTTTATGCTTTTAAAGCGGAACGCCTGCCGTTCGCGCATTG 60			
OY	21 ValleusSerSeCySGlyLyvSGLyGIyASPaaLaLAGLnGLyGINProalagLYArg 40			
Dd	61 GTACTGTCTCTTCGGCGrTAAGAAGCGGAGCGCCCAAGGCGCGCTGTGTCG 120			
OY	41 GlualaPcoAlaProValValGlValValtHnValHiSProGlnThrValaleutnr 60			
Dd	121 GAAGCCCCGCGCCCGCTCGTCGTgrCGrAACCGTCarCCGSAACCGTCGATTgACC 180			
OY	61 ValGiuleuProGLiyArgleugInuSERleaUrTHriAlaspVALarglaGlnvalgly 80			
Dd	181 GFCGAGTTSCCGGGGCGTTTGGMATCGTCGCrAACCGCGATGTCGCGCCAAGTCGGC 240			
OY	81 GlyllellieglnhyARgleuePheginGluNGlySetTyValARGalaglYGlnProeu 100			
Dd	241 GGcAtCaTCACAaaaAcCGCcTGTCCCAAAGAGcAgTTAgtccGTcccgCACgcggctg 300			
OY	101 TyrglinleAspSerseThrtYrglualasentleGuSerllaRgaRglaginleua 120			
Dd	301 TArcGaTrCAacAGttCCActTaRGaacMAatCTGSAAAAGcccGcGcCAacttgca 360			
OY	121 ThrAlaginalaThrleulaalySaLaaspaLaaspleuaLaagtYrlySpProeuVal 140			

Db	361	ACGGCTCAGGCAACGCTTGGCCAAAGCGGAGTCGGATTGTGGCGGATACCAAGCCTTGGTT	420
Qy	141	AlAlaAgluaIaValSerArgIngluTyraSpAlaAlaValThrAlaValArgSerAla	160
Db	421	GCCCCCAAGCCGTCACCCGCGAGAAATACGATGCTCGGTAAACGGCAACGTTCTGCC	480
Qy	161	GluuAaGlyVallyValaAlaGlnAlaAlaIleuysSerAlaGlyIleAenLeuAsnArg	180
Db	481	GAGGCAAGTGTCAAAAGACACAGCGCGCAATCAAAATCTCCGGCATTAATCTGAACCGT	540
Qy	181	SerArgIleThrAlaProIleSerGlyPheIleGlyInSerlyValSerGlyIleThr	200
Db	541	TCGGCGATTCCGGCGCGATTTCGGCTTATGGTCACTGATCCAAAGTTTCCGAAGTACG	600
Qy	201	LeuIeuAsnAaGlyAspThrThrValIleuAlaThrIleArgInThrAsnProMetTyr	220
Db	601	CTGTGAATCGGGCGGATACGACCGTCTGGCAACCATCCGCCAAACCAATCCGATGTAT	660
Qy	221	ValAsnValThrGlnSerAlaSerGluValMetIleuAsnArgInIleAaGluGly	240
Db	661	GTGAACCTTACCCAGTGTGATCCGAAGTATGAAATTTGGCCGTCAGATAGCGGAAGGC	720
Qy	241	IysIeuIeuAlaAlaAspGlyValIleAlaValGlyIleLysPheAspAspGlyThrVal	260
Db	721	AAATCTGCTGGCGCGGATGGTGATTCGGCTCGCATCAAAATTTGACACGGCACAGTT	780
Qy	261	TyrProGluIubArgIleArgLeuIeuPheAlaAspProValValAsnIuSerThrGlyIn	280
Db	781	TACCTGAAAAAGCCGCGCTGCTGTTGGCATCCGGTCTGTCAACGAATCGACCGGTACG	840
Qy	281	IleThrIeuArgAlaAlaValProAsnAspGlnAsnIleIeuMetProGlyIleuTyraI	300
Db	841	ATTACCTCTGGCGCGCGCTGACCGAAGATCAGAAATCTCTGATGCCCGTCTGTATGTG	900
Qy	301	ArgValIeuIeuAspGlnValAlaValAspAsnAlaPheValValProGlnGlnAlaVal	320
Db	901	CGCGTGTGATGGACCAAGTGGCGGTGATACCGCATTTGTGTGCCGACGACCGGCTGA	960
Qy	321	ThrArgIleValAlaIysAspThrValMetIleValAsnAlaGlnGlyIleMetGluProArg	340
Db	961	ACGGCGGGTGGAAAGATACCGTATGATTTGTGATGCCCAAGCGGTATGGAACCCCGC	1020
Qy	341	GluValThrValAlaGlnGlnGlnGlyThrAsnTrpIleValThrSerGlyIleuLysAsp	360
Db	1021	GAGGTAAACGGTGGCAACAGCAGGGGTACGAATGTGATTGTACGTGCGGTCTGAAGGAC	1080
Qy	361	GlyAspLysValValIleGluGlyIleSerIleAlaGlyIleThrGlyAlaLysVal	380
Db	1081	GGGGACAAAGGTGGTGTGGAAAGGATACGATACCGCGGTATTAACGGGTGGCAAAAGGTA	1140
Qy	381	ThrProLysGluTrpAlaSerSerGluAsnGlnAlaAlaIaIaProGlnSerGlyValGln	400
Db	1141	ACGCCCAAAAGATGGGCTGCTCTGAATAACCAAGCCGCCGCTCTCAATCCGGCTTCAG	1200
Qy	401	ThrAlaSerGluAlaLysThrAlaSerGluAlaGln	412
Db	1201	ACGGCATCTGAAGCCAAACTGCTTGAAGCGGAA	1236
RESULT 2			
ID	ABKS4080	ABKS4080 standard; DNA; 1239 BP.	
XX	ABKS4080;		
XX	AC		
DT	05-JUN-2002	(first entry)	
XX	DNA encoding Neisseria cell surface polypeptide #2.		
DE	Cell surface protein; gene; ds; antibacterial; antimicrobial.		
XX	Neisseria meningitidis.		
OS			
XX			
PN	WC200216612-A2.		

XX 28-FEB-2002.
 PD 21-AUG-2001; 2001WO-GB03759.
 PF 24-AUG-2000; 2000GB-0020952.
 PR (MICR-) MICROSCIENCE LTD.
 XX Lane JD, Hughes MJG, Santangelo JD;
 XX WPI; 2002-280941/32.
 DR P-PSDB; AAU91064.
 XX Novel peptide encoded by *Neisseria meningitidis*, useful for manufacture
 PT of medication for treatment or prevention of condition associated with
 PT infection by *Neisseria* or Gram-negative bacteria -
 XX
 XX Claim 1; Page 16-19; 79pp; English.
 XX
 XX The invention relates to polypeptides located on the cell surface of
 CC *Neisseria meningitidis*, and the polynucleotides encoding them. The
 CC sequences of the invention are useful for therapeutic or diagnostic use,
 CC in the manufacture of a medication for use in treatment or prevention of
 CC a condition associated with infection by *Neisseria* or Gram-negative
 CC bacteria. The sequences are also useful for screening potential
 CC antimicrobial drugs or for detection of virulence. Sequences
 CC ABK5479-ABK54095 represent *Neisseria meningitidis* polynucleotides of the
 CC invention.
 XX
 XX Sequence 1239 BP; 269 A; 337 C; 388 G; 245 T; 0 other;
 SQ
 Alignment Scores:
 Pred. No.: 4.44e-155 Length: 1239
 Score: 1989.00 Matches: 406
 Percent Similarity: 99.03% Conservative: 2
 Best Local Similarity: 98.54% Mismatches: 4
 Query Match: 98.51% Indels: 0
 DB: 24 Gaps: 0
 US-09-889-756a-2 (1-412) x ABK54080 (1-1239)

QY 161 GluAlaGlyValIySaIaAlaGlnAlaAlaIleYsSerAlaGlyIleAsnLeuAsnArg 180
 DB 481 GAGGCAAGCGCTTAAGCGCGGCGGCGGCAATCAATCCGCGGCGATCAGCTGACCGT 540
 QY 181 SerArgIleThrAlaProIleSerGlyPheIleGlyGlnSerIyValSerGlyIleThr 200
 DB 541 TCGGCATTAACCGCGCGGCGGCTTTCGCGCTTATCGGTACATCCAAAGTTTCGGAAGTACG 600
 QY 201 LeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAsnProMetTyr 220
 DB 601 TTGCTGAACCGCTGCGGATGCGACGCTACTGCGACCAATCCGCAAAATCCATGATAT 660
 QY 221 ValAsnValThrGlnSerAlaSerGlyValMetIleYsLeuArgArgGlnIleAlaGlyI 240
 DB 661 GTGACACTTACCCAGTGTGATCCGAAAGTATGAAATTCGCGCTCAGATACCGAAGGC 720
 QY 241 LysLeuLeuAlaAlaAspGlyValIleAlaValGlyIleYsPheAspAspGlyThrVal 260
 DB 721 AAACGTGCGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 780
 QY 261 TyrProGlnIySglYarGleuLeuPheAlaAspProValValAsnGlnSerThrGlyGln 280
 DB 781 TACCTGAAAGGCGCGCTGCTGTTGCGGATCCGCGCTCAACGAATGACCGGTACG 840
 QY 281 IleThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetProGlyLeuTyrVal 300
 DB 841 ATTAACCTGCG 900
 QY 301 ArgValLeuMetAspGlnValAlaValAspAsnAlaPheValValProGlnGlnAlaVal 320
 DB 901 CGCGTGTATGACCAAGTGGCGGTGATGATGATGATGATGATGATGATGATGATGATG 960
 QY 321 ThrArgGlyAlaIyAspThrValMetIleValAsnAlaGlnGlyIleMetGluProArg 340
 DB 961 ACCGCGCGTCCGAAAGTACCGATGATGATGATGATGATGATGATGATGATGATGATG 1020
 QY 341 GluValThrValAlaGlnGlnGlnIyIleThrAsnTrpIleValIleThrSerGlyLeuYsAsp 360
 DB 1021 GAGGTACCGGTGCGCAACGACGAGGTACGATGATGATGATGATGATGATGATGATG 1080
 QY 361 GlyAspIySglValValGlnGlyIleSerIleAlaGlyIleThrGlyAlaIySglVal 380
 DB 1081 GGGGACAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1140
 QY 381 ThrProIySglUtrPalSerSerGluAsnGlnAlaAlaProGlnSerGlyValGln 400
 DB 1141 ACCGCCAAAGATGGGCGTGTCTGAACCAAGCCGCGCTCAATCCGCGCTTACG 1200
 QY 401 ThrAlaSerGlnAlaIySglThrAlaSerGlnAlaGlu 412
 DB 1201 ACCGCACTGATGAACCAACCTGCTCTGAAGCGA 1236
 RESULT 3
 ID AAA81488 standard; DNA; 20352 BP.
 AC AAA81488;
 DT 04-DEC-2000 (first entry)
 XX
 DE N. meningitidis partial DNA sequence gnm_36 SEQ ID NO:36.
 XX
 KW *Neisseria meningitidis*; *Neisseria gonorrhoeae*; genome; immunogenic;
 KM antigen; vaccine; diagnosis; infection; antibacterial; identification;
 XX *Meningococcus B*; MenB; ds.
 OS *Neisseria meningitidis*.
 XX
 XX WO200022430-A2.
 XX
 XX 20-APR-2000.
 XX

PF 08-OCT-1999; 99W0-US23573.
 XX 09-OCT-1998; 98US-0103794.
 PR 30-APR-1999; 99US-0132068.
 XX (CHIR) CHIRON CORP.
 XX Frazer CM, Hickey E, Peterson J, Tettein H, Venter JC;
 PI Maigani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlatto V;
 PI Rappoli R, Pizzi M;
 XX WPI: 2000-318079/27.
 DR
 XX
 XX Isolated nucleotide sequences of *Neisseria meningitidis* which can be
 PT used in the diagnosis and treatment of *N. meningitidis* infection and
 PT other *Neisseria* infections, for example, *N. gonorrhoea*.
 XX
 XX
 PS Claim 7, Page 622-628; 1760pp; English.

XX The present invention describes methods of obtaining immunogenic
 CC proteins from *Neisseria* genomic sequences. AAA81453 to AAA82414
 CC represent specifically claimed *Neisseria meningitidis* genomic DNA
 CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
 CC *Neisseria* DNA sequences and their corresponding proteins; AAA81254 to
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
 CC isolation of *Neisseria meningitidis* DNA sequences; and AAA81322 to
 CC AAA81452 represent *Neisseria meningitidis* MenB polynucleotide ORF
 CC sequences, which are all used in the exemplification of the present
 CC invention. The nucleic acid sequences, protein sequences, and antibodies
 CC against them, can be used in the manufacture of a composition. The
 CC composition can be used as a medicament (or in the manufacture of a
 CC medicament) for treating, preventing or diagnosing infection due to
 CC *Neisseria* bacteria. For example, some of the identified proteins could
 CC be components of vaccines against *Meningococcus B*; against all serotypes;
 CC and/or against all pathogenic *Neisseriae*. Identification of sequences
 CC from the bacterium will also facilitate production of biological probes,
 CC particularly organism-specific probes. Attempts to make efficacious
 CC *Meningococcus B* vaccines have failed mainly due to antigen tolerance.
 CC Multivalent vaccines have also been tried but none have successfully
 CC overcome antigenic variability. The provision of further, complete
 CC sequences may provide an opportunity to identify secreted or surface
 CC exposed proteins that may be presumed targets for the immune system and
 CC which are not antigenically variable or at least more conserved than
 CC other more variable regions.

XX
 SO Sequence 20352 BP; 5082 A; 5105 C; 5364 G; 4800 T; 1 other;

Alignment Scores:

Pred. No.: 1,32e-153 Length: 20352
 Score: 1989.00 Matches: 406
 Percent Similarity: 99.03% Conservative: 2
 Best Local Similarity: 98.54% Mismatches: 4
 Query Match: 98.51% Indels: 0
 DB: 21 Gaps: 0

US-09-889-756a-2 (1-412) x AAA81488 (1-20352)

QY 1 MetAlaPheTyrAlaPheIyAlaMerAArgAlaAlaAlaLeuAlaAlaValAlaLeu 20
 DB 12358 ATGGCTTTTAAAGCTTTAAAGCGCATGCGCGCGCGGTGGCGCGCGGTGCATTG 12417
 QY 21 ValLeuSerSerCySgLyLySgLyGlyYAspAlaIaIngLyGlnProAlaGlyArg 40
 DB 12418 GTACTGTCGTCTTCGCGTAAAGCGGAGACCGCGGCGGCGGCGGCGGCGTGGTGG 12477
 QY 41 GluAlaProAlaProValAlaGlyValAlaThrValHisProGlnThrValAlaLeuThr 60
 DB 12478 GAAAGCCCTTGGCGCGCGTGTGTCTAAACCGTCCATCCGCAACCGTGCATTGACC 12537
 QY 61 ValGluLeuProGlyArgGluSerLeuArgThrAlaAspValArgAlaGlnValGly 80
 DB 12538 GTCGAGTTGGCGGCGGCGTTTGAATCGTGCATCCGCGCATGTCGCGCCCAAGTGGCG 12597

QY 81 GlyIleIleGlnIyAspGluPheIngLySerTyrValArgAlaGlyGlnProLeu 100
 DB 12598 GGCATCATCCAAAACCGCTGTTCCAAAGAGCAGTATATCGTGGCGGACCGCGCTG 12657
 QY 101 TyrGlnIleAspSerSerThrTyrGlnAlaAsnLeuGluSerLeaArgAlaGlnLeuAla 120
 DB 12658 TATCAGATCGACAGTTCCACTTATGAAGCAGGTCTGAAAAGCGCGCGCAACCTGCA 12717
 QY 121 ThrAlaGlnAlaThrIleAlaIyAspAlaAspAlaAspLeuAlaArgTyrLyProLeuVal 140
 DB 12718 ACGGCTACGCAACGCTTGCCTCAAGCGGATCGCGATTTGGCGCATCAACCTTGTT 12777
 QY 141 AlaAlaGlnAlaValSerArgIngLyTyrAspAlaAlaValThrAlaIyArgSerAla 160
 DB 12778 GCGCGCAACCGCTGACCGCGGAGAAATACGATGTCGGTAAACGCGGAACGTTGCGC 12837
 QY 161 GluAlaGlyValIyAspAlaIaIngAlaIleIySerAlaGlyIleAsnLeuAsnArg 180
 DB 12838 GAGCAGCGCTTAAAGCGCGCAGCGCATCAATCCGCGCATCCACCTGAACCGT 12897
 QY 181 SerArgIleThrAlaProIleSerGlyPheIleGlyGlnSerIyValSerGlyIyThr 200
 DB 12898 TCGCGCATTTACCGCGCGCATTTCCGCTTATCGTCAATCCAAAGTTCCGAAGTACG 12957
 QY 201 LeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAsnProMetTyr 220
 DB 12958 TTGCTGAACCTGGCGCATGCGACCGTACTGGCAGCATCCGCCAACAATCCGATGTAT 13017
 QY 221 ValAsnValThrGlnSerAlaSerGluValMetIyLeuArgArgGlnIleAlaGlnGly 240
 DB 13018 GTGAACCTTACCAAGTGTGATCCGAAGTATGAATGGCCGCTCAGATAGCGGAAGGC 13077
 QY 241 LysLeuLeuAlaIaAspGlyValIleAlaValGlyIleIySspPheAspAspGlyThrVal 260
 DB 13078 AACTGCTGCGCGCGCATGTTGATTCGCGTCCGATCAAAATTTGACGACGACAGTT 13137
 QY 261 TyrProGlnIySgLyArgLeuLeuPheAlaAspProValValAsnGlnSerThrGlyGln 280
 DB 13138 TACCTGAAAAGCGCGCGCTGTTGGCGCATCCGCGCTCAACGAATGACCGGTGACG 13197
 QY 281 IleThrLeuArgAlaAlaValProAsnAspGlnIleLeuMetProGlyLeuTyrVal 300
 DB 13198 ATTACCTGGCGCGCGCGTACGAAAGATCAGATCAATCTTGAATGCGCGGTCTGTATGTG 13257
 QY 301 ArgValLeuMetAspGlnValAlaValAspAsnAlaPheValValProGlnGlnAlaVal 320
 DB 13258 CGCGTCTGATGACCAAGTGGCGGTGATACGCAATTTGTTGGCGGACGACGCGGTA 13317
 QY 321 ThrArgGlyAlaIyAspThrThrValMetIleValAsnAlaGlnGlyIyMetGluProArg 340
 DB 13318 ACGGCGCGTCCGAAAGATACCGTATGATTTGTAATCCCAAGCGGTATAGGAACCCGCG 13377
 QY 341 GluValThrValAlaIngIngIngIyThrAsnThrIleValThrSerGlyLeuIyAsp 360
 DB 13378 GAGGTAAAGGTTCGCAACAGCAGGATGACGAATTTGATGTCGTGGTGTGAAGGAC 13437
 QY 361 GlyAspLysValValValGluGlyIleSerIleAlaGlyIleThrGlyAlaIyLysVal 380
 DB 13438 GGGGACAAGTGTGTGGGAAGCATCATGATCCGCGTAAACGCGTGCAGAAAAGTGA 13497
 QY 381 ThrProIySgIyTyrAlaSerSerGluAsnGlnAlaAlaProIySerGlyValGln 400
 DB 13498 ACGGCCAAGATGGGTGTCTGAAAACCAAGCCGCGCGCTCAATCCGCGGTTCAG 13557
 QY 401 ThrAlaSerGlnAlaIyThrAlaSerGlnAlaGln 412
 DB 13558 ACGGCATCTGAACCAACCTGCTTGAAGCGAAA 13593
 RESULT 4
 AAF21611/c
 ID AAF21611 standard; DNA; 349980 BP.
 XX
 AC AAF21611;

QY 381 ThrProlysgIuTPAlaSerSerGluAsnGlnAlaAlaProGlnSerGlyValGln 400
Db 295242 ACCGCCAAGATGGCGCTGCTGAAAAACCAAGCCCGCGCTCAATCCGGGCTTACG 295183
QY 401 ThrAlaSerGluAlaGlyThrAlaSerGluAlaGln 412
Db 295182 ACGGCATCTGAAGCCAAACCTGCTTGTGAAGCCAAA 295147
RESULT 5
ID AAA81489 standard; DNA; 837096 BP.
XX AAA81489;
AC
XX
XX
DT 04-DEC-2000 (first entry)
XX
XX N. meningitidis partial DNA sequence gnm_37 SEQ ID NO:37.
DE
XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
KM Meningococcus B; MenB; ds.
XX
XX Neisseria meningitidis.
OS
XX W0200022430-A2.
PN
XX 20-APR-2000.
PD
XX
PF 08-OCT-1999; 99WO-US23573.
XX
PR 09-OCT-1998; 98US-0103794.
PR 30-APR-1999; 99US-0132068.
XX
PA (CHIR) CHIRON CORP.
PI Frazer CM, Hickey E, Peterson J, Tettein H, Venter JC,
PI Maignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V,
PI Rappuoli R, Pizzi M;
XX
XX WPI; 2000-318079/27.
DR
PT Isolated nucleotide sequences of Neisseria meningitidis which can be
PT used in the diagnosis and treatment of N. meningitidis infection and
PT other Neisserial infections, for example, N.gonorrhoea -
XX
XX
XX Claim 7; Page 629-865; 1760pp; English.
XX
XX The present invention describes methods of obtaining immunogenic
XX proteins from Neisseria genomic sequences. AAA81453 to AAA82414
XX represent specifically claimed Neisseria meningitidis genomic DNA
XX sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
XX Neisseria DNA sequences and their corresponding proteins; AAA81254 to
XX AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
XX isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
XX AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
XX sequences, which are all used in the exemplification of the present
XX invention. The nucleic acid sequences, protein sequences, and antibodies
XX against them, can be used in the manufacture of a composition. The
XX composition can be used as a medicament (or in the manufacture of a
XX medicament) for treating, preventing or diagnosing infection due to
XX Neisserial bacteria. For example, some of the identified proteins could
XX be components of vaccines against Meningococcus B; against all serotypes;
XX and/or against all pathogenic Neisseriae. Identification of sequences
XX from the bacterium will also facilitate production of biological probes,
XX particularly organism-specific probes. Attempts to make efficacious
XX Meningococcus B vaccines have also been tried but none have successfully
XX Multivalent vaccines have also been tried but none have successfully
XX overcome antigenic variability. The provision of further, complete
XX sequences may provide an opportunity to identify secreted or surface
XX exposed proteins that may be presumed targets for the immune system and
XX which are not antigenically variable or at least more conserved than
XX other more variable regions.

SQ Sequence 837096 BP; 207534 A; 227065 C; 205215 G; 197280 T; 2 other;
Alignment Scores:
Pred. No.: 1,19e-151 Length: 837096
Score: 1988.00 Matches: 406
Percent Similarity: 99.03% Conservative: 2
Best Local Similarity: 98.54% Mismatches: 4
Query Match: 98.51% Indels: 0
DB: 21 Gaps: 0
US-09-889-756a-2 (1-412) x AAA81489 (1-837096)
QY 1 MetAlaPheTyrAlaPheLysAlaMetArgAlaAlaAlaLeuAlaAlaLeu 20
Db 361153 ATGGCTTTTATGCTTTTAAAGCGATGCTGGCGCCGCTTGCTCCGCTTCATTG 361094
QY 21 ValLeuSerSerGlyGlyGlyGlyAspAlaAlaGlnGlyGlnProAlaGlyArg 40
Db 361093 GTACTGTCGCTTGGCGGTAAAGCGAGACGCGCGGCGGCGGCGCTGCTGCGG 361034
QY 41 GluAlaProAlaProValAlaGlyValAlaThrValHisProGlnThrAlaLeuThr 60
Db 361033 GAAGCCCTCGCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360974
QY 61 ValGluLeuProGlyArgLeuGluSerLeuArgThrAlaAspValArgAlaGlnValGly 80
Db 360973 GTGAGTTGCGGGGCGCTTGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360914
QY 81 GlyIleIleGlnLysArgLeuPheGlnGlnGlySerTyrValArgAlaGlyGlnProLeu 100
Db 360913 GGCATCATCCAAAACCCCTGCTTCAAGAGGAGATATATCCGTCGCGACAGCCCTG 360854
QY 101 TyrGlnIleAspSerSerThrTyrGluAlaAsnLeuGluSerAlaArgAlaGlnLeuAla 120
Db 360853 TATCAGATTCAGACTTCACTTATGAGACAGGCTTGAAAGCCGCGCCCAATGCGCA 360794
QY 121 ThrAlaGlnAlaThrLeuAlaLysAlaAspAlaAspLeuAlaArgTyrLysProLeuVal 140
Db 360793 ACGGCTCAGGCAAGCTTGGCAAGCGGATGCGGATTTGGCGCGATFCAAGCTTTGGT 360734
QY 141 AlaAlaGluAlaValSerArgGlnGluTyrAspAlaAlaValThrAlaLysArgSerAla 160
Db 360733 GCGGCCCAACCGCTCAGCCGCGAGATACGATGCTGCGTAAACCGGAAACCTTCGCC 360674
QY 161 GluAlaGlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleLeuAsnArg 180
Db 360673 GAGCGAGCGCTTAAAGCGCGCGAGCGCGCAATCAATCCCGCGCATCACCTGAACCT 360614
QY 181 SerArgIleThrAlaProIleSerGlyPheIleGlyGlnSerLysValSerGluGlyThr 200
Db 360613 TCGGCGATTACCGGCGGATTTCCGGCTTATCGGTCAATCAAGTTTCCGAAGGTACG 360554
QY 201 LeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAsnProMetTyr 220
Db 360553 TTGGTGAACCTGCGCGATGCGACCGTACTGCGACCAATCCGCAACCAATCCGATGTAT 360494
QY 221 ValAsnValThrGlnSerAlaSerGluValMetLysLeuArgArgGlnIleAlaGluGly 240
Db 360493 GTGAACCTTACCCAGTGTGATCCGAGTGAATGAATTTGCGCGCTCGATAGCCGAAGGC 360434
QY 241 LysLeuLeuAlaAlaAspGlyValIleAlaValGlyIleLysPheAspAspGlyThrVal 260
Db 360433 AAACGTGTGCGCGGATGTGTGATTTGCGGTGCGCATCAATTTTGAACGACGACGTT 360374
QY 261 TyrProGluLysGlyArgLeuLeuPheAlaAspProValValLengLysThrGlyGln 280
Db 360373 TACCTGAAAAAAGCGCGCTGCTTGGCCGATCCGCGCTCAACGATGACGCGGTACG 360314
QY 281 IleThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetProGlyLeuVal 300
Db 360313 ATTACCTTGCGCGCGCGCGATCGAAGATCAAGATATCTTGATGCCCGGTCTGTATG 360254
QY 301 ArgValLeuMetAspGlnValAlaValAspAsnAlaPheValValProGlnGlnAlaVal 320

Db 360253 CGCGTCTGATGAGCAACAGTGGCGGTGATACGATTTGTGTGCGCCAGCGCGGTA 360194
 QY 321 ThrArgGlyAlaIleAspThrValMetIleValAsnAlaGlnGlyGlyMetGluProArg 340
 Db 360193 ACGGCGCGGTGCAAGAGATACCGTATGATTTGTGATGATGCGCAAGGGGATGAAACCCCGC 360134
 QY 341 GluValThrValAlaGlnGlnGlnGlyThrAsnTrpIleValIleuSerGlyLeuLysAsp 360
 Db 360133 GAGGTAAACGGTTGGCAACACAGGGTACGAATTGATTTACGTGGGTCTGAGAGAC 360074
 QY 361 GlyAspLysValIleValGluGlyIleSerIleAlaGlyIleThrGlyAlaLysLysVal 380
 Db 360073 GGGGACAAAGTGTGTGGAAGCATCATTCGCCGTATACGGGTGCGCAAAAAGGTA 360014
 QY 381 ThrProLysGluTrpAlaSerSerGluAsnGlnAlaIleAlaProGlnSerGlyValGln 400
 Db 360013 ACGCCCAAGATGGCGCTGCTGMAAACCAAGCGCGCCCTCAATCCGGCGTTGAG 359954
 QY 401 ThrAlaSerGluAlaLysThrAlaSerGluAlaGln 412
 Db 359953 ACGGCATCTGAAGCAACCTGCTTCTGAACGAAA 359918
 RESULT 6
 AB239062
 ID AB239062 standard; DNA; 1236 BP.
 XX AC AB239062;
 XX 07-MAR-2003 (first entry)
 XX N. gonorrhoeae nucleotide sequence SEQ ID 2713.
 XX Antibacterial; infection; vaccine; gene therapy; gene; ds.
 XX Neisseria gonorrhoeae.
 XX WO200279243-A2.
 XX 10-OCT-2002.
 XX 12-FEB-2002; 2002WO-IB02069.
 XX 12-FEB-2001; 2001GB-0003424.
 XX (CHIR-) CHIRON SPA.
 XX Fontana MR, Pizsa M, Maignani V, Monaci E;
 XX WPI; 2003-058415/05.
 XX P-PSDB; ABP78092.
 XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a
 XX medicament for treating or preventing N. gonorrhoeae infection -
 XX Disclosure; Page 386; 815pp; English.
 XX The present invention relates to proteins from Neisseria gonorrhoeae.
 XX Also disclosed are the nucleic acid molecules encoding the proteins and
 XX antibodies that specifically bind to the proteins. The composition
 XX comprising the protein, nucleic acid or antibody is useful for the
 XX manufacture of a medicament for treating or preventing N. gonorrhoeae
 XX infection, this may be in the form of a vaccine or gene therapy.
 XX Sequences given in records AB237706-AB242016 represent nucleic acid
 XX molecules of the invention.
 XX Sequence 1236 BP; 268 A; 334 C; 390 G; 244 T; 0 other;
 Alignment Scores:
 Pred. No. 2.78e-151 Length: 1236
 Score: 1943.00 Matches: 394
 Percent Similarity: 97.57% Conservative: 8
 Best Local Similarity: 95.63% Mismatches: 10

Query Match: 96.24% Indels: 0
 DB: 25 Gaps: 0
 US-09-889-756a-2 (1-412) x AB239062 (1-1236)
 QY 1 MetAlaPheTyrAlaPheLysAlaMetArgAlaAlaAlaLeuAlaIleValAlaLeu 20
 Db 1 ATGGCTTTTATGCTTCTTAAGCGCATCGTGGCGCCCGTGGCTGGCGCGTGGCATTTG 60
 QY 21 ValLeuSerSerGlyLysGlyValAspAlaAlaGlnGlyGlyProAlaGlyArg 40
 Db 61 GCACGTGCTTTCGGTAAAGCGGAGACGGGCGAGGGGGGAGCTGGCGGTCGG 120
 QY 41 GluAlaProAlaProValAlaGlyValValThrValHisProGlnTrpValAlaLeuThr 60
 Db 121 GAACCCCGCGCGCGCGTGGTGGGTGTAACCTTCATCCGCAACCGTGGCATTTAGC 180
 QY 61 ValGluLeuProGlyArgLeuGluSerLeuArgThrAlaAspValArgAlaGlnValGly 80
 Db 181 GTCGAGTTGGCGGGGCGTTTGAATCGCTGCTGCCCGCATGTCGCCCAAGTCGGC 240
 QY 81 GlyIleIleGlnLysArgLeuPheGlnGlySerTyrValArgAlaGlyGlnProLeu 100
 Db 241 GGCATCATCAAAAACCGCTGTCCAAAGAGCGATTATGTCCGCGCGGACCCGCTG 300
 QY 101 TyrGlnIleAspSerSerThrTyrGluAlaAsnLeuGluSerAlaArgAlaGlnLeuAla 120
 Db 301 TATCAGATCGACACTTCACTTATGAAGCAGTGTGAAAGCGCGCGCAACTGGA 360
 QY 121 ThrAlaGlnAlaThrLeuAlaLysAlaAspAlaAspLeuAlaArgTyrLysProLeuVal 140
 Db 361 ACGGCACAGCAACGCTTGCCAAAAGCGGATGGGATTGGCGGCTTCAAAACCGCTGTT 420
 QY 141 AlaAlaGluAlaValSerArgGlnGlyTyrAspAlaAlaValThrAlaLysSerAla 160
 Db 421 TCCGCCGATCCATCACTAAACAAAGATACGATCTCGGTAAACGGGAAACGTTCTGCC 480
 QY 161 GluAlaGlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAspArg 180
 Db 481 GAGCGCGCGCTTAAAGCGCGCGCGCGCATCAATCCCGCGCATCAATCAACCGT 540
 QY 181 SerArgIleThrAlaProIleSerGlyPheIleGlyGlnSerLysValSerGlnGlyThr 200
 Db 541 TCGGCATTAACGGCGCGGATTTCCGGCTTATCGGTACGTCAAAAGTTCCGAAGTAGC 600
 QY 201 LeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAsnProMetTyr 220
 Db 601 CTGTGAATCGGGCGGATACACTGTTTATAGCCACCATCCGCCAAACCAATCCGATGAT 660
 QY 221 ValAsnValThrGlnSerAlaSerGluValMetLysLeuArgArgGlnIleAlaGluGly 240
 Db 661 GTGACGTTACCCAGTGTGATCGAGTATGATGAATGCGCGCGCGCATGACCAAGGC 720
 QY 241 LysLeuLeuAlaAlaAspGlyValIleAlaValGlyIleLysPheAspAspGlyThrVal 260
 Db 721 AAGCTGTGGCGCGGATGTGTGCGATTCGGGTGCGCATCAATTGACGACGATGCGTT 780
 QY 261 TyrTrpGluLysGlyArgLeuLeuPheAlaAspProValValLeuGluSerThrGlyGln 280
 Db 781 TATCTGTAAGAAAAGCGCGCTGCTTGTGATCCGACCTTGAAGATGACGCGGTAG 840
 QY 281 IleThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetProGlyLeuTyrVal 300
 Db 841 ATTAACCTTGGCGCGCGCGATGATGAACAGATCAAGATATCTGTGATGCGCGGTATGTG 900
 QY 301 ArgValLeuMetAspGlnValAlaValAlaAspAsnAlaPheValValProGlnGlnAlaVal 320
 Db 901 CGTGTGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 960
 QY 321 ThrArgGlyAlaLysAspThrValMetIleValAsnAlaGlnGlyGlyMetGluProArg 340
 Db 961 ACGCGCGCGTGAAGATACCGGTGATGATGATGATGATGATGATGATGATGATGATGATG 1020

QY 341 GluValThrValAlaGlnGlnGlnGlnGlnThrSerTrpIleValThrSerGlyLeuLysAsp 360
DB 1021 GAGGTAAAGCTGGCGCAACAGAGGATCGAATTGATTGATCGGTCTGAAGAC 1080
QY 361 GlyAspLysValIleValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 380
DB 1081 GGGGACAGGCTGTTGGAGGACATCATCGCCGATGACGGGGCGAAAGGTA 1140
QY 381 ThrProLysGluTrpAlaSerSerGluAsnGlnAlaIleAlaProGlnSerGlyValGln 400
DB 1141 ACGGCTAAAGATAGGCGCCCTCTGAATAATCAAGCTGCGCCCTCAAGCGGCTTCAG 1200
QY 401 ThrAlaSerGluAlaLysThrAlaSerGluAlaGln 412
DB 1201 ACGGCATCTGAAGCCAACTGCTTCTGAAAGCAAA 1236
RESULT 7
AA211064
ID AA211064 standard; cDNA; 4879 BP.
AC AA211064;
XX
XX 01-NOV-1999 (first entry)
DE E. coli acra and acrb genes.
XX
XX Acra; acrb; tolC; organic solvent resistance; indigo production;
KM streptoid conversion; flooded crude oil treatment; ds.
XX
OS Escherichia coli.
XX
XX
FH Key Location/Qualifiers
FT CDS 329..1522
FT /tag= a
FT /product= acra
FT sig_peptide 329..400
FT /tag= b
FT mat_peptide 401..1519
FT /tag= c
FT CDS 1545..4694
FT /tag= d
FT /product= acrb
XX
XX JP11221080-A.
XX
PD 17-AUG-1999.
XX
PF 09-FEB-1998; 98JP-0027537.
XX
PR 09-FEB-1998; 98JP-0027537.
XX
PA (MEIJ) MEIJI SEIKA KAISHA LTD.
XX
XX WPI; 1999-520718/44.
DR P-PSDB; AAY32854; AAY32855.
XX
XX Escherichia coli having organic solvent resistance - useful for
PT production of indigo, conversion of steroid and treating flooded
PT crude oil
XX
XX Claim 2; Page 6-8; 14pp; Japanese.
XX
XX This sequence encodes the acra and acrb proteins. The invention relates
CC to a method for the preparation of E coli having organic solvent
CC resistance or of increased organic solvent resistance including a step of
CC transforming the E. coli with at least one gene of acra, acrb and tolC.
CC The E. coli can be used for the production of indigo from indol,
CC conversion of a steroid such as cholesterol and treatment of flooded
CC crude oil.
XX
SQ Sequence 4879 BP; 1157 A; 1215 C; 1343 G; 1164 T; 0 other;
Alignment Scores:

Pred. No.: 4,07e-53 Length: 4879
Score: 762.50 Matches: 168
Percent Similarity: 61.31% Conservative: 76
Best Local Similarity: 42.21% Mismatches: 129
Query Match: 37.77% Indels: 25
DB: 20 Gaps: 6
US-09-889-756a-2 (1-412) x AA211064 (1-4879)
QY 11 AlAlaIleAlaLeuAlaIleAlaValAlaLeuValLeuSerSerCysGlyLysGlyValAsp 30
DB 359 GCGGTCTTGTAGTCTCTCTCAGGACGTTAGCCCTTACAGAGAT--GACGCAAAACAG 415
QY 31 AlAlaGlnGlyGlnGlnProAlaGlyArgGluAlaProAlaProValGlyValVal 50
DB 416 GCCCAACAGAGTGGCCAG-----CAGATGCCCGCGGTGGGTAGTA 457
QY 51 ThrValHisProGlnThrValAlaLeuThrValGluLeuProGlyArgLeuGluSerLeu 70
DB 458 ACAGTCAAACTGAACCTCTGACAGATCAACCGAGCTTCGGGTCCACAGTGCCTAC 517
QY 71 ArgThrAlaAspValArgAlaGlnValGlyGlyIleIleGlnLysArgLeuPheGlnGlu 90
DB 518 CGATCCGACAGAGTTCTCTCTCTCAAGTTAGCGGATTTCTGAAAGCTAATTTCAAGAA 577
QY 91 GlySerTyrValArgAlaGlnGlnProLeuTyrGlnIleAspSerSerThrTyrGluAla 110
DB 578 GGTAGCGACATCGAAGAGAGTGTCTCTCTCAAGTATGATCTCGACCTATACAGCG 637
QY 111 AsnLeuGluSerAlaAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaValAsp 130
DB 638 ACATACGACAGATCGAAAGGTGATCTGCGAAAGCCAGGCTCCAGCCAAATATCGCAA 697
QY 131 AlaAspLeuAlaArgTyrLysProLeuValAlaAlaGlnAlaValSerArgGlnLysTyr 150
DB 698 TTGACGCTGAATGTTTCTTCAAGAACTCTCGTACTCAGTACATCAAGTACAGTAC 757
QY 151 AspAlaAlaValThrAlaLysArgSerAlaGlnAlaGlyValLysAlaAlaGlnAlaAla 170
DB 758 GATGAGCTCTGGGTGTATGCGCAACAGCGAAGTCTCGGTAACTGGCGGAAAGCTGCC 817
QY 171 IleLysSerAlaGlyIleAsnLeuAsnArgSerArgIleThrAlaProIleSerGlyPhe 190
DB 818 GTTGAACCTCGCGGAGTCAATCTGGCTTACCAAAATGCTCTCTCGATTAGGGTGC 877
QY 191 IleGlyGlnSerLysValSerGlnGlyThrLeuLeuAsnAlaGlyAspThrThrValLeu 210
DB 878 ATTGTAAGTCGAACGTACGGAAGGCGCATTTGATCAGAACGTCAGCGACTGCGCTG 937
QY 211 AlaThrIleArgGlnThrAsnProMetTyrValAsnValThrGlnSerAlaSerGluVal 230
DB 938 GCAACGCTGACGACATCTGATCCGATCTGATGATGACCCAGTCCACCAACGACTTC 997
QY 231 MetLysLeuArgArgGlnIleAlaGlnGlyLysLeuLeuAlaAlaAspGlyValIleAla 250
DB 998 CTGGCCTGAAACAGGAATCGGCAATGACACCTGAAACAGAGACGGCAAGCCCAA 1057
QY 251 ValGlyLysPheAspArgLysThrValTyrProGlyLysGlyArgLeuLeuPheAla 270
DB 1058 GTGTCACTGATCACCACTGACGCGCATTAAGTTCGCGAGACGCTACGCTGAATTCTT 1117
QY 271 AspProValValAsnGluSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAsp 290
DB 1118 GAGCTTACGTTATATACACACACTGGGTCTATACCTTACGCGCTATCTCCGAACCG 1177
QY 291 GlnAsnIleLeuMetProGlyLysTyrValArgValLeuMetAspGlnValAlaValAsp 310
DB 1178 GATCACACTGTGCTCGCGGTATGTTCTGTCGCGACGCTGAGAGAGAGGCTTAATCA 1237
QY 311 AsnAlaPheValValProGlnGlnAlaValThrArgGlyAlaLysAsp-----ThrVal 328
DB 1238 AACGTATTTTATGTCGCAACAGGGCGTAACTGATCGCGGTGGCATGCCACCGTA 1297

QY 329 MetIleValaIaenAlaGInGlyGlyMetCjUProArgIuValThrValAlaGInGIn 348
 Db 1298 CTGGTAGTGGCGGATGACAAAGTGAACCCGTCGATCGTCGACAGGCTATT 1357
 QY 349 GlyThrAntPrlIleValThrSerGlyLeuLeuAspGlyAspIysValValGlnGly 368
 Db 1358 GGCATTAAGTGGCTGGTACGAAAGTCTGAAGCAGGCGATCGGTATTAAGTGGG 1417
 QY 369 IleSerIleAlaGlyIleThrGlyAlaIysIysValThrPro 382
 Db 1418 CTG-----CAGAAAGTGGCTCGTGGTCTCCAGTTAAAGCA 1453
 QY 383 LysGluTrpAlaSerSerGlnAsnGlnAlaAlaProGlnSerGlyValGln 400
 Db 1454 CAAGAAATTACCGCTGATTAATACACGACGACCCGCA-----ACGGGTCTCAG 1501
 RESULT 8
 ABST8861
 ID ABST8861 standard; DNA; 7888 BP.
 XX
 AC ABST8861;
 XX
 DT 17-DEC-2002 (first entry)
 XX
 DE E. coli CFT073 genomic sequence #28.
 XX
 KM Pathogenic; Escherichia coli CFT073 infection; livestock; pyelonephritis;
 KM urinary tract infection; open reading frame; ORF; uropathogenic;
 KM antibacterial; atropathic; nephrotropic; gene; ds.
 XX
 OS Escherichia coli.
 XX
 PN WO200259320-A2.
 XX
 PD 01-AUG-2002.
 XX
 PF 19-OCT-2001; 2001WO-US46833.
 XX
 PR 19-OCT-2000; 2000US-242412P.
 XX
 PA (MISC) WISCONSIN ALUMNI RES FOUND.
 XX
 PI Blatner FR, Welch RA, Burland VD;
 XX
 WPI; 2002-691532/74.
 XX
 PT New DNA sequences of the pathogenic Escherichia coli CFT073 strain,
 PT useful for preventing or treating E. coli CFT073 infection in humans or
 PT livestock
 XX
 PS Claim 1; Page 120-124; 765pp; English.
 XX
 CC The present invention relates to polynucleotide sequences from the
 CC genome of the pathogenic Escherichia coli strain CFT073. Almost all
 CC the sequences present in E. coli CFT073 are absent in the previously
 CC sequenced laboratory strain K-12. The polynucleotide sequences of
 CC the invention are useful for preventing, diagnosing or treating
 CC E. coli CFT073 infection in humans or livestock. The polynucleotide
 CC sequences are useful for preventing urinary tract infections and
 CC pyelonephritis. Likewise, the polypeptides encoded by the different
 CC open reading frames (ORF1-5) are useful for generating a vaccine
 CC against uropathogenic E. coli strains. ABST8834-ABST9085 represent
 CC genomic sequences from E. coli strain CFT073.
 XX
 SQ Sequence 7888 BP; 1884 A; 1854 C; 2150 G; 2000 T; 0 other;
 Alignment Scores:
 Pred. No.: 7.89e-48 Length: 7888
 Score: 701.50 Matches: 162
 Percent Similarity: 60.63% Conserves: 69
 Best Local Similarity: 42.52% Mismatches: 133
 Query Match: 34.74% Indels: 17
 DB: 24 Gaps: 7

US-09-889-756A-2 (1-412) x ABST8861 (1-7888)
 QY 7 IysAlaIeArGlnAlaIleAlaIeAlaAlaValAlaIeUValIeSerSerCysGly 26
 Db 1064 AAAATATATGAAGTATATAGCAACATCTGTAGTGCAATGCTGCTTATCGGTTGT--- 1120
 QY 27 LysGlyGlyAspAlaIaGInGlyGlnProAlaGlyArgIuValaProAlaProVal 46
 Db 1121 -----GATATATACGAAAGTAAACATTCATCCCGTCAGAAACAGAA----- 1162
 QY 47 ValGlyValValThrValHisProGlnThrValAlaIeUValIeUProGlyArg 66
 Db 1163 GTGGCGCTTGTTCGCTCAATCTCAACCCGTTTCGGTATCTCATGTAAATTAACGGAGCT 1222
 QY 67 LeuGluSerLeuArgThrAlaAspValArgAlaGlnValGlyIleIleGlnIysArg 86
 Db 1223 ACCAGTCTCGCTCAGTGTCTGAAGTACGCCGAGTTGGGAAATTATTCAGAAACGC 1282
 QY 87 LeuPheGlnGluGlySerTyrrValArgAlaGlyGlnProLeuTyrrGlnIleAspSerSer 106
 Db 1283 TTAATTAAGAGAGGTATCTGTCTCAAGCTGGACAGCTCTTCAACGATTGATGCCGCC 1342
 QY 107 ThrTyrrGlnAlaAsnLeuGluSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeu 126
 Db 1343 AGTTATCAGGCTCATGTAAGTGAAGCCCGGACAGATTACACAGACAGACGCA---CTG 1399
 QY 127 AlaIysAlaAspAlaAspLeuAla---ArgTyrrLeuProLeuValAlaAlaGlnAlaVal 145
 Db 1400 GTAAAGCCGATGGCCAGAAAGCGCGTTATACCCGACTGTGGAAAGAAAGCGGTGT 1459
 QY 146 SerArgGlnGluTyrrAspAlaAlaValThrAlaIysSerAlaGlnAlaGlyValIys 165
 Db 1460 TCAACACAGAGTCTGATGATGCTCAGTCACTCACTGACCAACAAATAAAGCCAGTGAAG 1519
 QY 166 AlaAlaGlnAlaAlaIleIysSerAlaGlyIleAsnLeuAsnArgSerArgIleThrAla 185
 Db 1520 GCGAAAAAGCCGACCTGAAACCTGGCGCGCATTAATCTTACGACGACACGATTAACCGCA 1579
 QY 186 ProIleSerGlyPheIleGlyGlnSerIysValSerGlnGlyThrLeuLeuAlaGly 205
 Db 1580 CCGATTTGGGGGCTATTTGGCATTTGCTGCTGAACCCCTGGCCACTGAGTACCGCTCG 1639
 QY 206 AspThrThrValIleuAlaThrIleArgGlnThrAsnProMetTyrrValaIaenValThnGln 225
 Db 1640 CAGGATACACCGTTACGACTATTCGTCGTCGTGATCAATGATGTCGACTTCACTCCG 1699
 QY 226 SerAlaSerGluValMetIysLeuArgArgGlnIleAlaGluIlyIysLeuAlaAla 245
 Db 1700 TCCAGTGTGATTTATTAGCTGCGTAAACAG-----TCACGTGGGACCAAC 1747
 QY 246 AspGlyValIleAlaValGlyIleIysPheAspAspGlyThrValTyrrProGluIysGly 265
 Db 1748 AGTGACACCATGAGCGCTTCATCTTATTCGGAAGATGGCAACACTACACGAAAAAGGG 1807
 QY 266 ArgLeuLeuPheAlaAspProValValaIaenGluSerThrGlyGlnIleThrLeuArgAla 285
 Db 1808 CGCTGGAACCTACCCGAAAGTCGGGTGATGATGATCAACGTTCCGGTGAATTAACGGCA 1867
 QY 286 AlaValProAsnAspGlnAsnIleLeuMetProGlyLeuTyrrValaIaenValIleuMetAsp 305
 Db 1868 ATTTTCCCAATCCACACAGCAGTATTAACGGGAATGTTTTCGCGCTCGTGCAT 1927
 QY 306 GlnValAlaValaAspAsnAlaPheValValProGlnGlnAlaValaThrArgIyAlaIys 325
 Db 1928 GAAGGGGTATGTAAACGCTATTCTCGCGCTTCAACAGGCGCTTACCGCGATGCTTAA 1987
 QY 326 Asp-----ThrValMetIleValaIaenAlaGInGlyGlyMetCjUProArgIuValThr 343
 Db 1988 GGCATTAAGCACTCGCTGGTGTGAATTAAGACAAATTAAGTACGACGACGCAACCTCGAA 2047
 QY 344 ValAlaGlnGlnGlnGlyThrAntPrlIleValThrSerGlyLeuIysAspIysAspIys 363
 Db 344 ValAlaGlnGlnGlnGlyThrAntPrlIleValThrSerGlyLeuIysAspIysAspIys 363

Db 2048 ACGGAGAAAGCTATGATGATTAATGCTGCTGTAACGCGCTGCACACGAGCGA 2107
 QY 364 ValValValGluGlyIleSerIleAlaGlyIleThrGlyAlaValValThrProlys 383
 Db 2108 CTGATTTGTTGAGGT-----TCTGCCAAAGTCACTTCAAGCCACACCGCTCAAGCTT 2161
 QY 384 Glu 384
 Db 2162 GAA 2164
 RESULT 9
 ID AAF28550 standard; DNA; 99629 BP.
 XX AAF28550;
 AC AAF28550;
 XX
 DT 04-APR-2001 (first entry)
 XX
 DE Genomic fragment #37.
 XX
 KM Genomic library; bacteria; human upper airway; otitis media; sinusitis;
 KW bronchopulmonary; endocarditis; meningitis; ss.
 XX
 OS Moraxella catarrhalis.
 XX
 PN WO200078968-A2.
 XX
 PD 28-DEC-2000.
 XX
 PF 16-JUN-2000; 2000WO-US16649.
 XX
 PR 18-JUN-1999; 99US-0140121.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Lagace RE, Patterson C, Berg KL;
 XX
 DR WPI; 2001-041427/05.
 XX
 PT Genomic library for identifying diagnostic and therapeutic
 PT compositions; and for identifying virulence factors, regulatory
 PT elements and drug targets, comprises Moraxella catarrhalis nucleic
 PT acids -
 PS
 PS Claim 1; Page 391-415; 545pp; English.
 XX
 CC The present invention relates to a Moraxella catarrhalis genomic library
 CC comprising of a combination of 41 nucleic acid molecules (see
 CC AAF28514-AAF28554). The library has a number of uses described in the
 CC specification e.g. is useful for identifying diagnostic and therapeutic
 CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large
 CC aerobic, gram-negative diplococcus, normally found among the bacterial
 CC flora of human upper airways. M. catarrhalis is known to cause acute,
 CC localised infections such as otitis media, sinusitis and bronchopulmonary
 CC infection and life-threatening, systemic diseases including endocarditis
 CC and meningitis.
 XX
 SQ Sequence 99629 BP; 29233 A; 19222 C; 21909 G; 29264 T; 1 other;
 Alignment Scores:
 Pred. No.: 1.02e-40 Length: 99629
 Score: 631.50 Matches: 167
 Percent Similarity: 50.97% Conservative: 70
 Best Local Similarity: 35.91% Mismatches: 139
 Query Match: 31.28% Indels: 89
 DB: 22 Gaps: 10
 US-09-889-756a-2 (1-412) x AAF28550 (1-99629)
 QY 21 ValIeuSerSerGlyGlyValGlyValAlaAlaGlnGlyGlnProAlaGlyArg 40
 Db 1311 GTGCTGACGGCGTCACTCAACAGACGATACACG-----GTCCACCA 1355

QY 41 GluAlaProAlaProValValGlyValValThrValHisProGlnThrValAlaLeuThr 60
 Db 1356 CAATATGCCAATATCGACGATAGATGTTCAAACTATGCTGCAAACTGCGCCAGTATT 1415
 QY 61 ValGluLeuProGlyAlaGlyLeuGlySerLeuArgThrValAspValAlaGlnValGly 80
 Db 1416 CAAAGTTTTCAGAGCGGTGTGCGCATTGAAGCTTCAAGAGTTCGCCCGCAGTACG 1475
 QY 81 GlyIleIleGlnValArgLeuPheGlnGlyIleSerValArgAlaGlnProLeu 100
 Db 1476 GCGATTATGATGAAGTACTGTTCTAGAGAGTATGATGTCAGCAGCAGACCTTTA 1555
 QY 101 TyrglnIleAspSerThrTyr----- 108
 Db 1536 TATCGATTAATATGTCATAGTATACCGCAAGTGCACAGGTCGAAGCAGCGTTCAA 1595
 QY 109 -----GluAlaAsnLeuGluSerAlaArgAlaGlnLeu 119
 Db 1596 AACGCTGAGCGCGTGTGGTAAATGCCAGGCCACACAGCATTCAGTCAAGCAATTTA 1655
 QY 120 AlaThrAlaGlnAlaThrLeuAlaValAlaAspAlaAspLeuAlaArgTyrValProLeu 139
 Db 1656 GCGCTCAAGAGGCAATTTGGCTCAGCGCAGGCAAGATGTCATGCACTTGAGGCTTA 1715
 QY 140 ValAlaAlaGlnAlaValAspArgGlnGlyTyrAspAlaAlaValThrAlaValArgSer 159
 Db 1716 TTGGAGGTGAGCGATCTCTCGTCAAGCTTTGACCAACAGATTAACCGCATGCCAACC 1775
 QY 160 AlaGlnAlaGlyValValAlaAlaGlnAlaAla----- 170
 Db 1776 GCTGAAGCAATGTTTGGCTGCGCAGCGCAGTTCGCGCAGCAGATGTAATTAC 1835
 QY 171 -----IleValSerAlaGlyIleAsnLeuAsnArgSerArg----- 182
 Db 1836 AGTGCACACAGATCCGTCGCAAGTGCAGCGCCCAATTAACCAACGACGCTGATGCA 1885
 QY 183 -----IleThrAlaProIleSerGlyPheIleGlyIleSerValSerGlu 198
 Db 1896 GGTGTACATGATTCGCAAGGACCATCATGATGTTAAGCATCTCAGCGGTACAGCA 1955
 QY 199 GlyThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAspPro 218
 Db 1956 GGTGACTGTGATCCGCAATCAGCAACCAACATGTGTGCAATTAATCGTTGAGTAA 2015
 QY 219 MetTyrValAsnValThrGlnSerAlaSerGlyValMetLeuArgValGlnIleAla 238
 Db 2016 GTGATGTGACATCCAGCCAGTCTTCAACCAATGCTTAAGATTCGTGAACGATTCAG 2075
 QY 239 GluGlyValSerLeuLeuAlaAlaAspGlyValIleAlaValGlyIleValPheAspAspGly 258
 Db 2076 TCTGTCAAGTA-----GGTCAAGCGGTGCAGATGTTCACTGCGTGAAGATGCT 2129
 QY 259 ThrValTyrProGlnValGlyArgLeuPheAlaAspProValValAsnGluSerThr 278
 Db 2130 CAAACTATCAATCTTAGACACATTAATGCTTCCCATCAACAGATGATACAGCCAGC 2189
 QY 279 GlyGlnIleThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetProGlyLeu 298
 Db 2190 GGCACATGACCTTACCGCTGTGTTCCAAATCTCGATGTAAGCTTAATACCGGGATG 2249
 QY 299 TyrValArgValIleuMetAspGlnValAlaValAspAsnAlaPheValValProGlnGln 318
 Db 2250 TTTTGTATATCTGCTTAATTCAAAGTGTGTCATGACACAGCTTATATGCCACATCA 2309
 QY 319 AlaValThrArgGlyAlaValS-----AspThrValMetIleValAsnAlaGlnGly 336
 Db 2310 GCATCATGCGTACACCAAAAGCGCAAAACCAAGTTATATCTTAACGACGATTAACMA 2369
 QY 337 MetGluProArgGluValThrValAlaGlnGlnGlnGlyThrAsnThrPylIleValThrSer 356
 Db 2370 ATCGAGTTTGACGCGGTAACGACGAGTACTTATCATGATGTAATGATGTAACGAT 2429
 QY 357 GlyLeuValAspGlyValValValValGluGlyIleSerIleAlaGlyIleThrGly 376

DB 2098 GAAAACTGTCAGATTACCTCCGCTGGAGC 2130
RESULT 12
ID AAA65929 standard; DNA, 1395 BP.
XX AAA65929;
XX
XX 05-OCT-2000 (first entry)
XX
XX E. coli proliferation associated coding sequence SEQ ID NO:121.
XX
XX Escherichia coli; E. coli; proliferation; inhibition; screening;
XX KM antimicrobial; bacterial growth; antisense therapy; antibacterial; ds.
XX
XX Escherichia coli.
XX
XX WO200044906-A2.
XX
XX 03-AUG-2000.
XX
XX 27-JAN-2000; 2000WO-US02200.
XX
XX 27-JAN-1999; 99US-0117405.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Zyskind J, Ohlsen KL, Trawick J, Foreyth RA, Froelich JM, Carr GJ,
PI Yamamoto RT, Xu HH;
XX WPI; 2000-514822/46.
XX P-PSDB; AAB15924.
XX
XX Novel polynucleotides and polypeptides associated with microorganism
PT proliferation, used to identify inhibitors of bacterial growth and
PT proliferation, for use in antisense therapy -
XX
XX Claim 8; Page 129; 316pp; English.
XX
XX AAA65809 to AAA65889 and AAA66058 to AAA66138 represent nucleotide
CC sequences derived from Escherichia coli which inhibit E. coli
CC proliferation. AAA65890 to AAA66055 and AAB15886 to AAB16040 represent
CC nucleotide and protein sequences associated with E. coli proliferation.
CC AAA66056 and AAA66057 represent primers used for sequencing E. coli
CC proliferation inhibiting nucleotide inserts in an example from the
CC present invention. Methods from the present invention can be used to
CC identify a proliferation-regulated gene in a microorganism, by contacting
CC a microorganism with a proliferation-regulated gene activity inhibitory
CC nucleic acid identified in another organism, and determining if
CC inhibition occurs in the second microorganism. The nucleic acid sequences
CC identified as being required for bacterial growth and proliferation, can
CC be used for antisense therapy for killing bacteria.
XX
XX
SQ Sequence 1395 BP; 360 A; 372 C; 383 G; 280 T; 0 other;
Alignment Scores:
Pred. No.: 8.77e-27 Length: 1395
Score: 435.50 Matches: 124
Percent Similarity: 47.94% Conservative: 62
Best Local Similarity: 31.96% Mismatches: 176
Query Match: 21.57% Indels: 26
DB: 21 Gaps: 7
US-09-889-756A-2 (1-412) x AAA65929 (1-1395)
QY 22 LeuSerSerCysGlyLysGlyAspAlaAlaGlnGlyGly-GlnProAlaGly----- 39
DB 240 CTCCTCGAGTGCAGCCCGAGGGGCGAAGAAACAGCCAGCATCGCAGCGGTGTCTCG 239
QY 40 -----ArgGlnAlaProAlaProAlaValGlyValValThyValHisProGlnTh 56
DB 300 ACGGTATGATCGTTCCGCGCATTTAGCCCGCGTTCAAGCGCGACCGCGTAGAACAGGC 359

QY 56 rValAlaLeuThrValGlnLeuProGlyArgLeuGlnSerLeuArgThrAlaAspValAr 76
DB 360 AGTTCCGCGTTACTCTACCCGGGCTTGGCACCATTAACCGCGCTTAATACCGTTACGGTGGC 419
QY 76 gAlaGlnValGlyGlyLeileGlnLysArgLeuPheGlnGlySerThrValArgAl 96
DB 420 CACCGCGGTGGAGCGCAATGATGAGCTTACATTTCCAGAGAGCGCAGCGTCAAGC 479
QY 96 agLysInProLeuThrGlnIleAspSerSerThrThrGlnAlaAsnLeuGlnSerAlaAr 116
DB 480 AGGCGATTACTGCGCAAAATGATGACCCCGCTTCAAACTTGATTCATTAGCAAGGCCA 539
QY 116 gAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaLysAlaAspAlaAspLeuAlaArgTy 136
DB 540 GGGCGAACTGGCAAAAGATTAACCGCTTGGCAACGCCCGCGTACCTGGCGCGTTA 599
QY 136 rLysProLeuValAlaAlaGlnAlaValSerArgGlnGlyTyraAspAlaAlaValThrAl 156
DB 600 TCACAACCTGGCAAAACCAATCTCGTTCCCGCAGAGCTGATGCCCAAGCGGCT 659
QY 156 aLysArgSerAlaGlnAlaGlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGly 176
DB 660 GGTCACTGAACCGAAGGACCATTAAGCTGATGAAGCAAGCGTCCGCGCAGCT 719
QY 176 eAsnLeuAsnArgSerArgIleThrAlaProIleSerGlyPheIleGlyInserLysVa 196
DB 720 GCAACTCGACTGAGACCGCGATTACCGCACAGTCAGTGTGGGTGTCTCAAGCAGCT 779
QY 196 lSerGlnGlyThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnTh 216
DB 780 TGATGTTGTAACCAATCTCCAGTGTGATACACCGGAGTGTGTGATACCCGAGC 839
QY 216 rAsnProMetThrValAsnValThrGlnSerAlaSerGlnValMetLysLeuArgArgG 236
DB 840 GCATCTCATGATTGATTTACTTACCTGCGGAAGAGATATGCTACCTATGTCAGGC 899
QY 236 nIleAlaGlnGlyLysLeuLeuAlaAlaAspGlyValIleAlaValGlyLe--LysPh 255
DB 900 GCAGAAAGCGGAAACCGCTGTGTGAAGCGCTGGATCGCACCAACCTCGAAGAAAT 959
QY 255 eaAspAspGlyThrValTyreProGlnLysGlyArgLeuPheAlaAspProValValAs 275
DB 960 AAGTGAAGGACG-----CTGTTAAGCTCTGATTAACCAATCGA 998
QY 275 nGlnSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeuMe 295
DB 999 TGCCACTTACCGTATGATTAAGTGAAGCAGCTTTAATATCAGAGTATCGCTGTT 1058
QY 295 tProGlyLeuThrVal-----ArgValLeuMetAspGlnValAlaValAspAsnAlaPh 313
DB 1059 TCCCAATCAGTTGTTTAACGGCGCATGTTAGTCGAC-----ACCGAACAACCGCCT 1112
QY 313 eValValProGlnGlnAlaValThrArgGlyAlaLys--AspThrValMetIleValAs 332
DB 1113 AGTATGCCAACAGCGCCCTCGCAATGGGCATTAAGGCCATTTGTCGTGGTCTGAA 1172
QY 332 nAlaGlnGlyValMetGlnProArgGlnValThrValAlaGlnGlnGlyThrAsnTr 352
DB 1173 TAGCGAAACAGAGTGCAGCAACATCTGTGAGCGCGGCGCATTCAGACAGTCAAGAAAGT 1232
QY 352 pIleValThrSerGlyLeuLysAspGlyAspLysValValAlaGlnGlyLe----- 369
DB 1233 GGTGATCCGTGCGAGTATTTCTCGCGGCATTCGCTGTGACAGACGCGCATTTATCGCCT 1292
QY 370 -----SerIleAlaGlyIleThrGlyAlaLysValThrProLysGlnTr 385
DB 1293 GACCGAAGGGCGCAAGTGAAGTGTGAAGCCAGAGCGGCACATTAATCCGGAAGAGAA 1352
QY 385 pAlaSerSerGlnAsnGlnAla 392
DB 1353 AGCCACACGCGCGAATACGCG 1374

RESULT 13
 ID AAH84642 standard; DNA, 1395 BP.
 AAH84642
 AC
 XX
 XX
 DT 26-SEP-2001 (first entry)
 XX
 XX
 DE E. coli growth and proliferation related coding sequence SEQ ID NO:270.
 XX
 XX
 KM Escherichia coli; growth; proliferation; microbial; antimicrobial;
 XX bacterial infection; microorganism; ds.
 OS Escherichia coli.
 XX
 XX
 PN W0200134810-A2.
 XX
 PD 17-MAY-2001.
 XX
 PF 09-NOV-2000; 2000WO-US30950.
 XX
 PR 09-NOV-1999; 99US-0164415.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Forsyth RA, Ohlsen K, Zyskind J;
 DR WPJ; 2001-335933/35.
 DR P-PSDB; AAG98971.
 XX
 XX
 PT Novel nucleic acids that inhibit Escherichia coli proliferation, useful
 PT for screening for homologous genes and for designing expression vectors
 PT -
 XX
 XX
 PS Claim 9; Page 337-339; 522pp; English.
 XX
 XX
 CC AAH84373 to AAH84499 represent Escherichia coli growth and proliferation
 CC related DNA sequences (I). AAH84500 to AAH84670 encode the E. coli
 CC growth and proliferation related proteins given in AAG99078 and AAG98830
 CC to AAG98999. (I) can be used as potential targets for the generation of
 CC new antimicrobial agents, and for identification of compounds which
 CC interact with the gene products of (I). In addition the expression of
 CC (I) and the purification of the proteins, the purified proteins can be
 CC used to generate reagents and screen small molecule libraries or other
 CC candidate compound libraries for compounds that can be further developed
 CC to yield novel antimicrobial compounds. In addition, nucleic acid probes
 CC complementary to (I) that are specific for particular species of
 CC microorganisms can be used to identify particular microorganism species
 CC in clinical specimens, therefore, providing a rapid and dependable
 CC method by which to identify the causative agents of a bacterial
 CC infection. Also, antibodies generated against proteins translated from
 CC mRNA transcribed from proliferation-regulated sequences can also be used
 CC to screen for specific microorganisms that produce such proteins in a
 CC species-specific manner. AAH84371 and AAH84670 represent sequencing
 CC primers used in the isolation of E. coli growth and proliferation
 CC related sequence, which are used in an example from the present
 CC invention.
 CC
 CC
 SQ Sequence 1395 BP; 360 A; 372 C; 383 G; 280 T; 0 other;
 Alignment Scores:
 Pred. No.: 8.77e-27 Length: 1395
 Score: 435.50 Matches: 124
 Percent Similarity: 47.94% Conservative: 62
 Best Local Similarity: 31.96% Mismatches: 176
 Query Match: 21.57% Indels: 26
 DB: 22 Gaps: 7
 US-09-889-756A-2 (1-412) x AAH84642 (1-1395)
 22 leuSERserCysGlyIysGlyIysADPAlaIaGlnGlyIys-GlnProAlaGly----- 39
 240 CTCCCGAGATGTCACCCACAGGGCGAGCAAAACAAGCGCAGCAATGCGCCAGCGGATGTCG 299

QY	40	-----ArglnuAlaProAlaProValGlyValValThrValHisProGlnTh	56
DB	300	ACGRTGTAAGCGTTCGGCCCACTTAAGCCCGGTTCAAGCGCGCAGCCGCCGTGAACACAGC	359
QY	56	ValAlaLeuThrValGluLeuProGlyValGluGluSerLeuThrValAspValAr	76
DB	360	AGTTCGGGTACCTCACCGCGCTTGGGACCATTAACCGCCCTATACCGTTACGGTGGC	419
QY	76	GAlaGlnValAlaGlyLeuIleGlnLysArgLysPheGlnGlnGlySerThrValArgAl	96
DB	420	CAGCCGCTGGACGGCCCACTGATAGCGTTACATTTCCAGGAAGGCCAGCAGGTCAAGC	479
QY	96	AGLYGlnProLeuThrGlnIleAspSerSerThrThrLysLysLeuGluSerAlaAr	116
DB	480	AGGGGATTACTGGGAGAAATTGACCCGACGAGTTCMAAGTTGCATTAGCAACAGCCCA	539
QY	116	GAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaLysAlaAspLeuAlaArgTy	136
DB	540	GGGCCAACTGGCCAAAGATAAAGCAGCGTTGCCAAGCCCGCGTGCCTGGCGCGTTA	599
QY	136	FLYSerProLeuValAlaAlaGlnAlaValSerArgGlnGluThrAspAlaAlaValThrAl	156
DB	600	TCAACAACTGGCAAAACGAATCTGCTTTCCCGCCAGAGCTGATGCCCAACAGCGCT	659
QY	156	AlaYArgSerAlaGlnAlaGlyAlaLysAlaAlaGlnAlaLysSerAlaGlyTl	176
DB	660	GATCAGTGAACCAAGCAGCACCATTTAAGCGTGAAGAAGCAAGCGTTCAGCGCGCACT	719
QY	176	eAsnLeuAsnArgSerArgIleMetAlaProIleSerGlyPheIleGlnSerLysVal	196
DB	720	GCAATCTCAGCTGGAGCGGATTTACCGACACGTCAGTGTGCGGTGCTCTCAAGCAGT	779
QY	196	ISerGlnGlyThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnTh	216
DB	780	TGATGTTGTAAACCAATCTCCAGTGGTGAATACACCGGAGTCTGGTGATCACCCAGAC	839
QY	216	AsnProMetLysrValaAsnValThrGlnSerAlaSerGluValMetLysLeuAlaArgGl	236
DB	840	GCATCTTAATCATTTAGTCTTTACCCTGCCGGAAGCATATCCCTACCGTAGTGCAGC	899
QY	236	nIleAlaGlnGlyLysLeuLeuAlaAlaAspGlyValIleAlaValGlyLeu---LysPh	255
DB	900	GCAGAAACCCGGAAGAACCCGCTGCTGTGAAGAACCTCGGATGCCACCACTCGAAGAAATT	959
QY	255	eAspAspGlyThrValTyProGluLysGlyArgLeuLeuPheAlaAspProValAlaAs	275
DB	960	AAGTCAAGGACG-----CTGTAACTAGTAAACCAATCA-----	998
QY	275	nGluSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeuMe	295
DB	999	TGCCACTACCGGTACGATTAAGTAAGAAAGCAGCTTATATATACAGATGATGCGCTTT	1058
QY	295	cProGlyLeuTyVal-----ArgValLeuMetAspGlnValAlaValaAspAsnAlaPh	313
DB	1059	TCCCAATCAGTTGTGTAACGCGCGCATGTTATGTAGTAC-----ACCGAAACAAACGCGCT	1112
QY	313	eValValProGlnGlnAlaValAlaThrArgGlyAlaLys---AspThrValMetIleValAs	332
DB	1113	AGTATCCCAACAGCCCGCTTGCCAATGGGCAATGAAGAGCCATTGTTCTGGGTGGTGA	1172
QY	332	nAlaGlnGlyLysMetGlyProArgGluValThrValAlaGlnGlnGlnGlyThrAsnTr	352
DB	1173	TAGCGAAACAAAGTCAGCAAAACATCTGTGTATCCCGGGCATTCAGACAGTCAAGAACT	1232
QY	352	pIleValThrSerGlyLeuLysAspGlyAspLysValValaGluGlyTle-----	369
DB	1233	GTGTATCCGTGACAGATATTTCTGCGGGCGATCGCGTGTGACAAACGCGCATTTAGTCCT	1292
QY	370	-----SerIleAlaGlyIleMetnGlyAlaLysLysValThrProLysGluTr	385
DB	1293	GACCGAAGGGCGAAGAGTGTGGGAAGCCAGAGCCCACTACCTCCGGAAGAA	1352

XX Nucleotide sequence of PA14 mexa and mexB.
 DE Human pathogen; virulence polypeptide; virulence factor;
 KW pathogenic infection; Pseudomonas aeruginosa infection; ss.
 XX Pseudomonas aeruginosa.
 OS
 PN WC0927129-A1.
 XX
 PD 03-JUN-1999.
 XX
 PF 25-NOV-1998; 98WO-US25247.
 XX
 PR 25-NOV-1997; 97US-0066517.
 XX
 PA (GEHO) GEN HOSPITAL CORP.
 XX
 PI Ausubel F, Cao H, Drenkard E, Goodman HM, Mahajan-Miklos S;
 PI Rahme LG, Tan M, Tsongalis J;
 XX
 DR WPI; 1999-357851/30.
 XX
 PT Virulence factors useful in developing disease treatments
 PS Disclosure; Fig 14B; 228pp; English.
 XX
 CC The present sequence represents a Pseudomonas aeruginosa nucleic acid
 CC sequence. P. aeruginosa is an opportunistic human pathogen present in
 CC soil water and plants. The specification describes virulence polypeptides
 CC and nucleic acid sequences encoding such polypeptides. These sequences
 CC can be used to identify a compound which is capable of decreasing the
 CC expression of a pathogenic virulence factor. Compounds that inhibit
 CC the expression or activity of virulence factor polypeptides can be
 CC used to treat pathogenic infections, especially where the infection
 CC is a P. aeruginosa infection.
 CC note: the sequences given in the specification were poorly legible, and
 CC in some instances assumptions were made as to the identity of the base;
 CC it is therefore possible that the sequence given below is not entirely
 CC correct.
 CC
 SQ Sequence 1008 BP; 205 A; 331 C; 320 G; 152 T; 0 other;
 XX
 XX
 Alignment Scores:
 Pred. No.: 1,37e-21 Length: 1008
 Score: 370.50 Matches: 80
 Percent Similarity: 58.29% Conservative: 43
 Best Local Similarity: 37.91% Mismatches: 73
 Query Match: 18.35% Indels: 15
 DB: 20 Gaps: 5
 US-09-889-756A-2 (1-412) x AAX98220 (1-1008)
 QY 205 G1YAspThrThrValLeuAlaThrIleArgInlThrsAspMetTyValAsnValThr 224
 Db 1 GGCAGAGCAAAACGGATGCGACCGTCGACAGCTCGACCGGATCTAGTGTGACGTCAC 60
 QY 225 GlnserIAspSerGluValMetLysLeuArgArgInlIleAlaGluGlyLysLeu---Leu 243
 Db 61 CAGCGCTCCACCGCGCTGTCGTCATGCGCGCAACGCGCACGCGGCACTTGAGAGCGC 120
 QY 244 AlaIAspGlyValIleAlaValGlyIleLysPheAspAspGlyThrValTyProGlu 263
 Db 121 GCCCGGCAACACGCTGCGAAGGTCCTCTGAAGCTGGAGGCGGTAGCCAAATACCGCTG 180
 QY 264 LysGlyArgLeuLeuPheAlaAspProValValAsnGlnserThrGlyGlnIleThrLeu 283
 Db 181 GAAGCGCGCTCGAATCTCCGAGGTTCCGTCAGCAGACGACCGGCTCGTCACCATC 240
 QY 284 ArgAlaAlaValProAsnAspGlnAsnIleLeuMetProGlyLeuTyValArgValLeu 303
 Db 241 CGCGCGCTGTTCCCAACCCGCAACAGAGCTGCTGCCCGCATGTTCTGTCACGCGCAG 300

QY 304 MetAspGlnValAlaValAspAsnAlaPheValValProGlnGlnAlaValThrArgGly 323
 Db 301 TTGAGAGGAGCGCTCAAGCAGAAAGCATCTGCTCCGCAAGAGCGTGAACCCGCGAC 360
 QY 324 AlaLysAsp-----ThrValMetIleValAsnAlaGlnGlyMetGluProArgGlu 341
 Db 361 CTCAGAGGCCAGGCTACCGGCTGCTGTGTAACCGCGCAACAGAGCTGAGCTCGGGTG 420
 QY 342 ValThrValAlaGlnGlnGlnGlnIleThrAsnTrpIleValThrSerGlyLeuLysAspGly 361
 Db 421 ATCAAGGCCACCGCGGATCGCGCGACAGTGGCTGTCAACGAGCTGAACCGCGGC 480
 QY 362 AspLysValValValGluGlyIleSerIleAla-----GlyIleThrGlyAlaLysLys 379
 Db 481 GACAAAGTCAATACCGAAGCGCTGCAAGTTCGTGACAGCGGGGTGCGAG---GTCAAGACC 537
 QY 380 ValThrProLysGluTrpAlaSerSerGluAsnGlnAlaIleAlaProGlnSerGlyVal 399
 Db 538 GTGCGCGCGAAGATGTGCGGTCGCGCAGAGGCCGACCGCTCGC----- 585
 QY 400 GlnThrAlaSerGluAlaLysThrAlaSerGlu 410
 Db 586 -----GCCAAACCGACAGCAG 603
 RESULT 16
 AAS79919
 ID AAS79919 standard; cDNA; 2112 BP.
 XX
 AC AAS79919;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #15723.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSBQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR P-PSDB; ABG15732.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 1; SEQ ID No 15723; 103bp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations in
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at http://wipo.int/pub/published_pct_sequences.

XX Sequence 2112 BP, 547 A, 601 C, 532 G, 432 T, 0 other;

Alignment Scores:

Pred. No.:	1,34e-18	Length:	2112
Score:	339.00	Matches:	105
Percent Similarity:	46.56%	Conservative:	78
Best Local Similarity:	26.72%	Mismatches:	134
Query Match:	23	Indels:	77
DB:		Gaps:	11

US-09-889-756a-2 (1-412) x AAS79919 (1-2112)

QY 6 PheValAlaMetArgAlaAlaAlaLeuAlaAlaAlaValAlaLeuValLeuSerSerCys 25
 DB 947 TTCAGGAACCTGCCCTGACGCGCATATCACCCCTACCTTATGATATATACACACATG 1006
 QY 26 GlyValSerGlyGlyAspAlaAlaGlnGlyGlyProAlaGlyArgGlyAlaPro----- 43
 DB 1007 -----ACTTACCTCGCAACGCAAGATGCCACCGGCA 1039
 QY 44 -----AlaProValValGlyValValThrValHis-----ProGlnThrVal--- 57
 DB 1040 GCCGGAAACCATCATACCTGACGCGCTTACGCTCAGTTCGAGCTTCACTGACGCTTCCG 1099
 QY 58 -----AlaLeuThrValGlyLeuProGlyArgLeuGlnSerLeuArgThrAla 73
 DB 1100 GTTATGTAACCCGCTATGACATGACCGTTT-GGTACAGGTGACGCGGTACACCGCGAG 1158
 QY 74 AspValArgAlaGlnValGlyGlyIleIleGlnValArgLeuPheGlnGlnGlySerTyr 93
 DB 1159 GAA-----GGGCTGAGTCAATGACCGC-----GCATAC 1185
 QY 94 ValArgAlaGlyGlnProLeuTyrGlnIleAspSerSerThrTyrGlnAlaAsnLeuGlu 113
 DB 1186 GACAGCGCTGACGCTTATGCGGTGAAGACACGACGCGCATGAAGCGGTATGAA 1245
 QY 114 SerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaValAspAla----- 131
 DB 1246 TACAAACCGCGCGACATACGACAGTGCAGAAAGTGAATGTCGAAAGCCCGACGTGACGC 1305
 QY 132 -----AspLeuAlaArgTyrLeuProLeuValAlaAlaGlnAlaValSer 146
 DB 1306 AATATCGCGCAATTCAGCGTGAATCGTTATCGAAGATGCTCGGTACTCGTATCATCAGT 1365
 QY 147 ArgGlnGlnTyrAspAlaAlaValAlaThrAlaGlySerSerAlaGlnAlaGlyValValAla 166
 DB 1366 AAGCAAGATGACGATGAGCTGCTGATGCGCAACAGGGAATGCTGCGGTACTGCT 1425
 QY 167 AlaGlnAlaAlaIleLeuSerArgIleYIleAsnLeuAsnArgSerArgIleThrAlaPro 186
 DB 1426 GCGAAAGCTGCGGTGAACCTGCGACATCATCTGCTTACACCAAG-----GCACCT 1479
 QY 187 IleSerGlyPheIleGlyGlnSerLeuValSerGlnGlyThrLeuLeuAsnAlaGlyAsp 206
 DB 1480 CTCGATTACCGGTCGATTCGACACGGAAGCGCATGTCGTAAGACGCTCAG 1539
 QY 207 ThrThrValLeuAlaThrIleArgGlnThrAspProMetTyrValAsnValThnGlnSer 226
 DB 1540 GCGACCTCGCTGGACCGTGCACCACTTGATCGATCGATGATGATGATGATGATGATGATGAT 1599
 QY 227 AlSerGlnValMetIleValArgArgGlnIleAlaGlnGlyIleLeuLeuAlaAlaAsp 246
 DB -----

DB 1600 AGCAACACTTCTCGCCCTGGAACAGAACTGGCGAATGGCACGCTGAACAAAGAGAAC 1659
 QY 247 GlyValIleAlaValGlyIleLeuPheAspAspGlyThrValTyrProGlnIleGlyArg 266
 DB 1660 TGCAAAGCCAAAGTGTACATGATACACAGTACGCGCATTAAGTTCCTCCGACGACGTAAC 1719
 QY 267 LeuLeuPheAlaAspProValAlaAsnGlnSerThrGlyGlnIleThrLeuArgAlaAla 286
 DB 1720 CTGGAATTCCTGACGTTACCGTTGATCAGACCACTGGGTCTATCACCTTACCGGCTATTC 1779
 QY 287 ValProAsnAspGlnAsnIleLeuMetProGlyLeuTyrValArgValLeuMetAspGln 306
 DB 1780 TTCGCCAAGCCGATCACTCTGCTCCGCGC----- 1812
 QY 307 ValAlaValAspAsnAlaPheValValProGlnGlnAlaValThrArgGlyAlaLeuAsp 326
 DB 1812 ----- 1812
 QY 327 ThrValMetIleValAsnAlaGlnGlyIleMetGluPro-----ArgGluValThr 343
 DB 1813 AATATCGACGAATGATTAAGTCTGCGGCTTGGACCTTCAATCACCGCATCCCACTG 1872
 QY 344 ValAlaGlnGlnGlnGlyThrAsnTyrIleValThrSerGlyLeuValAsp----- 360
 DB 1873 ATCATTTAGTACAGATCGGACCGCTTGGTGTGTCGAAACATGATGACGACACACATATCT 1932
 QY 361 GlyAspIleValValGlnGlyIleSerIleAlaGly 373
 DB 1933 GCGGACAGACCGCTGATTCGCGAAGTACGCTTACAGA 1971
 DB
 RESULT 17
 ID AAS81885 standard; cDNA, 2112 BP.
 XX AAS81885;
 AC AAS81885;
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #17689.
 XX
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSB-) HYSBQ INC.
 XX
 PI Dmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR P-PSDB; ABG17698.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 1; SEQ ID No 17689; 103bp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2112 BP: 547 A; 601 C; 532 G; 432 T; 0 other;

Alignment Scores:

Pred. No.:	1,34e-18	Length:	2112
Score:	339.00	Matches:	105
Percent Similarity:	46.56%	Conservative:	78
Best Local Similarity:	26.72%	Mismatches:	134
Query Match:	16.79%	Indels:	77
		Gaps:	11

US-09-889-756A-2 (1-412) X AAS81865 (1-2112)

QY 6 PheIysAlaMetArgAlaAlaAlaLeuAlaAlaAlaValAlaLeuValLeuSerSerCys 25
 Db 947 TTCGAGAACTGCCCTGACCGCATATCACCCCTGATGATATATCCACACAGTG 1006
 QY 26 GlyIysGlyGlyAspAlaAlaGlnGlyGlnProAlaGlyArgGlnAlaPro----- 43
 Db 1007 -----ACTTACCTCGGACGACGAGATGCCACCGCA 1039
 QY 44 -----AlaProValValGlyValValThrValHis-----ProGlnThrVal--- 57
 Db 1040 GCCGGAACACCATGACCGTGGAGCGGTACGCTGACGTTCACTGACCTGTTCCG 1099
 QY 58 -----AlaLeuThrValGlnLeuProGlyArgLeuGlnSerLeuArgThrAla 73
 Db 1100 GTTATGTAACCGCTATGACCATGACCGTTT-GGTACAGTGAGCGCGGTACCGCGAG 1158
 QY 74 AspAlaArgAlaGlnValGlyGlyLeuLeuGlnArgLeuPheGlnGlnGlySerTyr 93
 Db 1159 GAA-----GGGCTGAGTCAATGACG-----GCATAC 1185
 QY 94 ValArgAlaGlyGlnProLeuTyrGlnLeuSerSerThrTyrGlnAlaAsnLeuGln 113
 Db 1186 GACGACCGGTGACGATTAATGCGGTGAAGACGCGGCGCATGAACCGGTATGAA 1245
 QY 114 SerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaValAlaAspAla----- 131
 Db 1246 TACAACCGCGCGGACATACGACGAAAGTGTCTTCAAAAGCCCGACGTGACGCC 1305
 QY 132 -----AspLeuAlaArgTyrTyrProLeuValAlaAlaGlnAlaValSer 146
 Db 1306 AATATCGCGCAATTGACGCTAATCGTATCAGAAACGCTCGGTACTCACTACATACAGT 1365
 QY 147 ArgGlnGlnTyrAspAlaAlaValThrAlaLysArgSerThrAlaGlnAlaValValSerAla 166
 Db 1366 AAGCAAGAGTACGATGAGCTCTGCGTGAATGCGCAACGCGAATGCTCGCGTATACGTCT 1425
 QY 167 AlaGlnAlaAlaLeuSerSerAlaGlyLeuLeuLeuAsnArgSerArgLeuThrAlaPro 186
 Db 1426 GCGAAAGCTGCGTGAAGTGCACAGATCAATCTGTTACACCAAG-----GCACCT 1479
 QY 187 IleSerGlyPheIleGlyGlnSerLysValSerGlnGlyThrLeuLeuAsnAlaGlyAsp 206
 Db 1480 CTCGATTAAGCGTCGATTCGAAAGTGAAGGCGCATTTGTGACAGAAAGGTCAG 1539

QY 207 ThrThrValLeuAlaThrIleArgGlnThrAsnProMetTyrValaValaValThrGlnSer 226
 Db 1540 GCGATCGCTGGCAACCGGACGAACTTGATCCGATTCATGATGATGACCCAGTCC 1599
 QY 227 AlaSerGlnValMetLysLeuArgArgGlnIleAlaGlnGlyLysLeuLeuAlaAlaAsp 246
 Db 1600 AGCAACACACTCTGCGCTGAAACAGAACTGGCGAATGGACGCTGAAACAGAGAC 1659
 QY 247 GlyValIleAlaValGlyIleLysPheAspAspGlyThrValTyrProGlnLysGlyArg 266
 Db 1660 TGCAAACCAAGTGTACTGATCACCAGTACGCGCTTAAGTCCCGCAGACGATACG 1719
 QY 267 LeuLeuPheAlaAspProValValaAsnGlnSerThrGlyGlnIleThrLeuArgAlaAla 286
 Db 1720 CTGGAATTCCTGACGTTACCGTTCATGACAGACCTGGTGTATACCCATCCGCGTATC 1779
 QY 287 ValProAsnAspGlnAsnIleLeuMetProGlyLeuTyrValArgValLeuMetAspGln 306
 Db 1780 TTCGGAACCGGATCACACTCTGCTGCGCGG----- 1812
 QY 307 ValAlaValaAspAsnAlaPheValValProGlnGlnAlaValaThrArgGlyAlaLysAsp 326
 Db 1812 ----- 1812
 QY 327 ThrValMetIleValaAsnAlaGlnGlyMetGlnPro-----ArgGlnValThr 343
 Db 1813 AATATCAGCAATGAGATTAAGTCTGCGGCTTGAGACTTTCATACACCGCATCCACTG 1872
 QY 344 ValAlaGlnGlnGlnGlyThrAsnThrIleValThrSerGlyLeuLysAsp----- 360
 Db 1873 ATCATTAAGTATGATCGGACCGCTGGTGTGCAACATGATGACGAAACCAATACT 1932
 QY 361 GlyAspLysValValaGlnGlyIleSerIleAlaGly 373
 Db 1933 GCGGCGACGACCTGATTCGCGAAGTACGCTTTCAGA 1971
 Db
 RESULT 18
 AAS82449
 ID AAS82449 standard; cDNA, 2310 BP.
 XX
 AC AAS82449;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #18253.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYS-E-) HYSBQ INC.
 XX
 PI Dermanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR P-PSDB; ABG18262.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 1; SEQ ID No 18253; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPRO at ftp.wipro.int/pub/published_pct_sequences.

XX Sequence 2310 BP; 562 A; 600 C; 634 G; 514 T; 0 other;

Alignment Scores:

Pred. No.:	2,58e-17	Length:	2310
Score:	324.00	Matches:	86
Percent Similarity:	48.62%	Conservative:	37
Best Local Similarity:	33.9%	Mismatches:	105
Query Match:	16.05%	Indels:	25
DB:	23	Gaps:	2

US-09-889-756a-2 (1-412) x AAS82449 (1-2310)

QY 22 LeuSerSerCysGlyLysGlyValAspAlaIaIaGlnGlyGly-GlnProAlaGly----- 39
DB 1527 CTCCCGAGTGCACGCCCGGCGGCGAGCAACAGCGCAGCATCGCAGCGGCGTGTGCG 1586
QY 40 -----ArgGluAlaProAlaProAlaValGlyValValHisProGln 56
DB 1587 ACGTGTATGCGTCCGCGCCATTAGCCCGGTTCAGCGCGGCGAGCCCGCTGTAACAGGC 1646
QY 56 rValAlaLeuThrValAlaGluLeuProGlyArgLeuGluSerLeuArgThrAlaAspValAr 76
DB 1647 AGTTCGCCGCTTACTCACCAGCGGCTTGCGACCATTAACGCCCTTAATACGTTAGCGTGG 1706
QY 76 gAlaGlnValGlyGlyIleIleGlnLysArgLeuPheGlnGlnGlySerTyrValArgAl 96
DB 1707 CAGCGCGTGCAGCGCACTGATGAGCTTATCATTTCCAGGAGGCGCAGCGTCAAGC 1766
QY 96 aglyGlnProLeuTyrGlnIleAspSerSerThrTyrGlnAlaAsnLeuGluSerAlaAr 116
DB 1767 AGGGGATTTACTGGCAAAATATGACCCAGCGCACTTAAAGTTGATGCAAGGCCA 1826
QY 116 gAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaValAlaAspAlaAspLeuAlaArgTyr 136
DB 1827 GGGCACTCGCAAAATATAGCCACGCTTGCCCAAGCGCCGCTAACCTGCGCGCTTA 1886
QY 136 rLysProLeuValAlaIaIaGlnAlaValSerArgGlnGlnTyrAspAlaIaValThrAl 156
DB 1887 TCACAACACTGGCAAAACCAATCTCGTTCCCGCAGAGATGATGCCCAAGCGGCT 1946
QY 156 alyArgSerAlaGlnAlaGlyValLysValAlaGlnAlaIaIaLysSerAlaGlyTl 176
DB 1947 GGTGATGAACCCGAGGACCACTTAAGCTGATGAAGCAAGCGCTTCCAGCGGCGAGCT 2006
QY 176 eAsnLeuAsnArgSerArgIleThrAlaProIleSerGlyPheIleGlyGlnSerLysVa 196
DB 2007 GCAACTGACTGGAGCGCGATTAACGCAACGATGAGTGTGCTGTCTCAAGCAGGT 2066
QY 196 lSerGlnGlyThrLeuAsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnTh 216

DB 2067 TGATGTTGTTACCAAACTTCAGTGTGATATACCAACCGGATCGTGTGATACCCAGAC 2126
QY 216 rAsnProMetTyrValAsnValThrGlnSerAlaSerGluValMetLysLeu----- 233
DB 2127 GCATCTATGATGATTATGCTTTACCTCGCGGAAGGATATCGCTACCGTAGGCAAGC 2186
QY 234 -----ArgArgGlnI 237
DB 2187 GCAGAAAGCCGGAACCCGCTGTGTGTAAGAGCTGGAGTCGACCAACTGCAAGAAATT 2246
QY 237 eAlaGlnGlyLysLeuLeuAlaIaAspGlyValIle 249
DB 2247 AAGTGAAGCAACGCTGTTAAGTTAAGTAAACCAATA 2283

RESULT 19

ID ABZ40695 standard; DNA, 1176 BP.
AC ABZ40695;
XX 07-MAR-2003 (first entry)
DT 07-MAR-2003 (first entry)
XX N. gonorrhoeae nucleotide sequence SEQ ID 5979.

XX Antibacterial; infection; vaccine; gene therapy; gene; ds.

XX Neisseria gonorrhoeae.

XX W0200279243-A2.

XX 10-OCT-2002.

XX 12-FEB-2002; 2002WO-IB02069.

XX 12-FEB-2001; 2001GB-0003424.

XX (CHIR-) CHIRON SPA.

XX Fontana MR, Piazza M, Maignani V, Monaci E;

XX WPI, 2003-058415/05.

XX P-PSDB; ABP79725.

XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a medicament for treating or preventing N. gonorrhoeae infection -

XX Disclosure; Page 624; 815pp; English.

XX The present invention relates to proteins from Neisseria gonorrhoeae. CC Also disclosed are the nucleic acid molecules encoding the proteins and CC antibodies that specifically bind to the proteins. The composition CC comprising the protein, nucleic acid or antibody is useful for the CC manufacture of a medicament for treating or preventing N. gonorrhoeae CC infection, this may be in the form of a vaccine or gene therapy. CC Sequences given in records ABZ37706-ABZ42016 represent nucleic acid CC molecules of the invention.

XX Sequence 1176 BP; 301 A; 275 C; 379 G; 221 T; 0 other;

Alignment Scores:

Pred. No.:	9.22e-17	Length:	1176
Score:	313.00	Matches:	116
Percent Similarity:	43.57%	Conservative:	67
Best Local Similarity:	27.62%	Mismatches:	176
Query Match:	15.50%	Indels:	62
DB:	25	Gaps:	11

US-09-889-756a-2 (1-412) x ABZ40695 (1-1176)

QY 7 LysAlaMetArgAlaIaIaLeuAlaIaIaValAlaValLeuSerSerCysGly 26
DB 7 AAATATGTAAGGAGCGCTGTGCGCGTGTGCGCGCGCA----- 48

```

Qy 27 LysGlyIAspAlaIleGlyGlyGlnProAlaGlyArgGluAlaProAlaProVal 46
Db -----GGGGTTTGGGGCGGAGTGTCTTATCTGAAGCCGAGCCGAGCTCT 96
Qy 47 ValGlyValValThrValHisProGlnThrValAlaLeuThrValGluLeuProGlyArg 66
Db 7ATATTACGAAACGTCAGCGCGGCGGATACAGCCGCGAGCTTCCGCAAGCGGCGAG 156
Qy 67 LeuGluSerLeuArgThrAlaAspValArgAlaGlnValGlyGlyIleIleGlnIleArg 86
Db 157 ATTTCGCGGTCGCAACCTGTATCGCTGCGCGCGAGCTTGGGCGACATTAAGAGCTT 216
Qy 87 LeuPheGlnGlySerThrValArgAlaGlyGlnProLeuThrGlnIleAspSerSer 106
Db 217 TATGTCAAACTCGGCGCAACAGTCTCAAAAAGGCGGATTTGGCGGAATCAATTCGACC 276
Qy 107 ThrTyr -----GluAlaAsnLeuGluSerAlaArgAlaGlnLeu 119
Db 277 ACGCAGACCAACAGCAGTATGGAATAATCCAAATTGGAACGCTATCAGCGCAAGCTG 336
Qy 120 AlaThrAlaGlnAlaThrLeuAlaIleAlaAspAlaAspLeuAlaArgTyrLeuProLeu 139
Db 337 GTCTCGGACAGATTCGATTCGGCGAGCGCGAGAAAGAAATTAACGCTCAGCGCGCTTG 396
Qy 140 ValAlaAlaGluAlaValSerArgGlnGluTyrAspAlaAlaValThrAlaIleAspSer 159
Db 397 TGAAGGATGATGCGACCTTAAGAAAGATTTGGAACCGCGCAGATGCCCTTGCCGCC 456
Qy 160 AlaGluAlaGlyValIleValAlaAlaGlnAlaIleIleSerAlaGlyIleAsnLeuAsn 179
Db 457 GCCAAGCCCAATGTTGCGAGTGAAGCTTAAATCAACAGCAACAAATTTCCATCAT 516
Qy 180 Arg -----SerArgIleThrAlaProIleSerGlyPheIleGly 192
Db 517 ACCGCCAGATGCGATTTGGGCTACACGCGCATTCACGCGAGATGAGACGCGAGTGTG 576
Qy 193 GlnSerIleValSerGlyGlyThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThr 212
Db 577 GCGATTCCTCGTGAAGAGGCGCAGACTGTGAACCGCGCG ----- 615
Qy 213 IleArgGlnThrAsnProMetTyrValAsnValThrGlnSerAlaSerGluValMetIle 232
Db 616 -----CAGTCTACGCGCAGCATTTGCCAATTTGGCGAAT -----CTGAGATATATGTTG 663
Qy 233 LeuArgArgGlnIleAlaGluGlyLysLeu -----LeuAlaIleAspGlyValIleAla 250
Db 664 AACAAATGCAATTCGCGAGCGCGATATTACCAAGGAGGAAGCGCGCGCAGGATTTTCG 723
Qy 251 ValGlyIleLysPheAspAspGlyThrValTyrProGluIleGlyArgLeuLeuPheAla 270
Db 724 TTTACGATTTTGTCCGAACCGGATACG -----CCGATTAGCGCAAGCTCGACAGCGTC 777
Qy 271 AspProValVal -----AsnGluSerThrGlnGlnIleThr 282
Db 778 GACCCCGGCTGACACGAGTGTCTCGGCGGCTACCAACACAGATACGATACGCGCTTC 837
Qy 283 -----LeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetPro 296
Db 838 AATCGCGCTATTTATATGCTTCCTGTTGTGCGCAATCCGCGACGCGCAACCTGCCACG 897
Qy 297 GlyLeuTyrValArgValLeuMetAspGlnValAlaValAspAsnAlaPheValValPro 316
Db 898 GGGATGACGACGAAATACGCTGGAATACGACGCTGGAATAATGTTGCTTATTCG 957
Qy 317 GlnGlnAlaVal ---ThrArgGlyAlaIleAspThrValMetIleValAsnAlaGlnGly 335
Db 958 TCGCTGACCGTGAATAATCGCGCGCAAGCGCTTCGACCGCTGTGGCGGCGACGCGC 1017
Qy 336 GlyMetGluProArgGluValThrValAlaGlnGlnGlnGlyThrAsnTyrIleValThr 355
Db 1018 AAGGACGTGGAACCGCAATTCGACCGGCTATGAAAGACATATGAATACGAGTGA 1077
Qy 356 SerGlyLeuIleAspGlyAspGlyValValValGlnGlyIleSerIleAlaGlyIleThr 375

```

```

Db 1078 AGCGGTTGAAAGAGGGGCGCAAGTGTGTCATCTCCGAATATACCGCGCGA-GCAGCA 1136
Qy 376 GlyAlaIleLysValThrProLysGluTyrAlaSerSerGluAsnGlnAlaIlePro 395
Db 1137 GGAAGCGGGAACGCGCCCTTAGG-----CGGCCGCGCGCCG 1175

RESULT 20
AA121306
ID AA121306 standard; DNA; 1179 BP.
XX
AC AA121306;
XX
XX 08-OCT-1999 (first entry)
DT
XX
DE Neisseria gonorrhoeae complete ORF85 sequence.
XX
KM Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW treatment; Neisseria infection; meningitis; septicemia; gonorrhea; ss.
XX
OS Neisseria gonorrhoeae.
XX
EN W09924578-A2.
XX
PD 20-MAY-1999.
XX
PF 09-OCT-1998; 98MO-IB01665.
XX
PR 01-SEP-1998; 98GB-0019016.
PR 06-NOV-1997; 97GB-0023516.
PR 14-NOV-1997; 97GB-0024190.
PR 18-NOV-1997; 97GB-0024386.
PR 27-NOV-1997; 97GB-0025158.
PR 10-DEC-1997; 97GB-0026147.
PR 14-JAN-1998; 98GB-0000759.
XX
PA (CHIR-) CHIRON SPA.
XX
PI Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;
XX
DR WPI: 1999-327407/27.
XX
DR P-PSDB; AAY38884.
XX
PT Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
XX diagnosis, treatment and prevention of infection
XX
PS Claim 9; Page 430; 524pp; English.
XX
CC Nucleotide sequences AA211972-Z12358 represent open reading frames
CC (ORFs) of Neisseria meningitidis and N. gonorrhoeae which encode
CC antigenic proteins (see AAY38499-Y38944). The antigenic proteins, their
CC fragments, their nucleic acids and antibodies are used for diagnosis,
CC prevention (as vaccines) or treatment of Neisseria infections,
CC such as meningitis, septicemia and gonorrhea. Both organisms
CC are closely related. Fragments of the nucleic acids are useful
CC as hybridisation probes and antisense reagents.
XX
SO Sequence 1179 BP; 304 A; 275 C; 378 G; 222 T; 0 other;

Alignment Scores:
Pred. No. Length: 1179
Score: 310.00 Matches: 115
Percent Similarity: 43.57% Conservative: 68
Best Local Similarity: 27.38% Mismatches: 176
Query Match: 15.35% Indels: 62
DB: 20 Gaps: 11

US-09-889-756a-2 (1-412) x AA121306 (1-1179)
Qy 7 LysAlaMetArgAlaAlaAlaLeuAlaAlaValAlaLeuValLeuSerSerCysGly 26
Db 7 AAATGATGAAATGGCGGCTGTGGCGCGCTCGCGCGCAACGGT----- 54

```

```

QY 27 LysGlyIAspAlaAlaGlnGlyGlnProAlaGlyArgGluAlaProAlaProVal 46
Db 55 -----TGGGGCGGATGGTCTTATCTGAAGCCGGAACCCAGCGTCT 96
QY 47 ValGlyValValThrValHisProGlnThrValAlaLeuThrValGluLeuProGlyArg 66
Db 97 TATATTACGAACCGTACGGCGCGCGATATACGCCGAGGTTCCGCGACGGCGCAG 156
QY 67 LeuGluSerLeuArgThrAlaAspValArgAlaGlnValGlyIleIleGlnLysArg 86
Db 157 ATTTCGCCGTCACACCTGTATCGTCCGCCGCGAGCTTCGGGCGAGATTAAAGCTT 216
QY 87 LeuPheGlnGluGlySerThrValArgAlaGlnProLeuTyrgLinnLeuAspSerSer 106
Db 217 TATGTCAAACTCGGCGCAACAGGTCAAAAGCGCGATTGTGGGAAATCAATTCGACC 276
QY 107 ThrTyr-----GluAlaLeuGluSerAlaArgAlaGlnLeu 119
Db 277 AGCGACACCAACACGATGATATGAAATAATTCAAATTTGAAACGTATCAGCGCAAGCTG 336
QY 120 AlaThrAlaGlnAlaThrLeuAlaLysAlaAspLeuAlaArgTyrLysProLeu 139
Db 337 GTGTCCGACAGATTGATTTGGCAGCGCGGAAAGAAATTAACCTCAGCGCGCTTG 396
QY 140 ValAlaAlaGluAlaValSerArgGlnGluTyrAspAlaAlaValThrAlaLysArgSer 159
Db 397 TGGAAAGATGATCGACCTCTAAAGAGATTGGAAAGCCGCGAGATCCGCTTGCGGCC 456
QY 160 AlaGluAlaGlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsn 179
Db 457 GCCAAAGCAATGTTCGCCAGTGAAGCTTAAATCAGACAGCAAAATTTCCATCAT 516
QY 180 Arg-----SerArgIleThrAlaProIleSerGlyPheIleGly 192
Db 517 ACCGCCAGTCCGATTTGGGCTACACGGCTTACCGCGACGATGACGCGACCGGTGTG 576
QY 193 GlnSerLysValSerGlyIleThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThr 212
Db 577 GCGATTCCTCCGTGAAGAGGGGCAAGCTGTGAACCGCGG----- 615
QY 213 IleArgGlnThrAsnProMetTyrValAsnValThrGlnSerAlaSerGluValMetLys 232
Db 616 -----CAGTCTACGCGACGATTTGCCAATTTGCGCAT-----CTGCAATATATGTTG 663
QY 233 LeuArgArgGlnIleAlaGlnGlyLysLeu-----LeuAlaAlaAspGlyValIleAla 250
Db 664 AACCAAAATGCAATTTGCCAGGCGCGATTTACCAAGTGAAAGCGCGCGAGATATTTCG 723
QY 251 ValGlyIleLysPheAspAspGlyThrValTyrProGluLysGlyArgLeuPheAla 270
Db 724 TTTACGATTTTGTCCGAACCGGATACG-----CCGATTAAAGCGGAAGCTCGACAGGCTC 777
QY 271 AspProValVal-----AsnGluSerThrGlyGlnIleThr 282
Db 778 GACCCCGCGGTGACACGATGCTCGGGCGGCTACCAACAGCACTACGATACGCGCTTCC 837
QY 283 -----LeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetPro 296
Db 838 AATCGCGCTATTATTATGCCCCGTTCTGTTGTGCCGAATCCGACGGCAAACTGCCACG 897
QY 297 GlyLeuTyrValArgValLeuMetAspGlnValAlaValAspAsnAlaPheValPro 316
Db 898 GGGATGACGACGCGAATACGGTTGAATCGACGCTGAAATAATGTTGCTTATTCCG 957
QY 317 GlnGlnAlaVal---ThrArgGlyAlaLysAspThrValMetIleValAsnAlaGlnGly 335
Db 958 TCCGTCACCGTGAATAATCCCGCGGAGCGCTTCGTACGCGTGTGGGTGGCGAGCGC 1017
QY 336 GlyMetGluProArgGluValThrValAlaGlnGlnGlyThrAsnTyrIleValThr 355
Db 1018 AAGGACGTGAACCGCAAAATCCGACCGGTATGAAACAGCTATGATCCGAAGTAAA 1077
QY 356 SerGlyLeuLysAspGlyAspLysValValGlnGlyIleSerIleAlaGlyIleThr 375

```

```

Db 1078 AGCGGTTGAAAGAGGGGGGACAAAGTGTCTATCTCCGAAATTAACCGCGCGCA-GCAGCA 1136
QY 376 GlyAlaLysLysValThrProLysGluThrPalasSerSerGluAlaGlnAlaAlaPro 395
Db 1137 GGAAGGCGCGCAACGCCCTTAGG-----CGGCCGCGCGCGCG 1175

RESULT 21
ID AAF21607 standard; DNA; 349980 BP.
XX
AC AAF21607;
XX
DE 13-MAR-2001 (first entry)
XX
KW Neisseria meningitidis B nucleotide sequence SEQ ID NO:108.
XX
FM Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
KM diagnosis; antigen; detection; infection; gene therapy; antibacterial;
ds.
XX
OS Neisseria meningitidis.
XX
EN W020006791-A1.
XX
PD 09-NOV-2000.
XX
PE 08-MAR-2000; 2000MO-US05928.
XX
PR 30-APR-1999; 99US-0132068.
XX
PR 08-OCT-1999; 99MO-US23573.
XX
PR 28-FEB-2000; 2000GB-0004695.
XX
PA (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMI.C RES.
XX
PI Piazza M, Hickey E, Peterson J, Tettein H, Venter JC, Masignani V,
PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R,
PI Frazer CM, Grandi G;
XX
DR WPI; 2000-647603/62.
XX
PT Neisseria meningitidis B full length genome sequence and open reading
PT frames are used to detect, treat and prevent Neisserial infections -
XX
PS Claim 7; Appendix A; 692pp; English.
XX
CC The present invention describes the full length genome of
CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
CC to AAF21613 represent fragments of the NMB genomic sequence, as the
CC sequence was too long to go in a record on its own it was split into 8
CC sequences which overlap each other at the beginning and end of each
CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at
CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
CC Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
CC AAF21606 represent PCR primers which are used in the exemplification of
CC the present invention. The NMB genome and fragments from it have
CC antibacterial activity, and can be used in vaccines and gene therapy.
CC Neisseria nucleic acids, proteins and/or antibodies which binds to the
CC proteins can be used in compositions for treating or preventing infection
CC due to Neisserial bacteria or as a diagnostic reagent for detecting the
CC presence of Neisserial bacteria or of antibodies raised to Neisserial
CC bacteria. Computers, computer memory, computer storage medium or computer
CC databases can be used in a search to identify open reading frames (ORFs)
CC or coding sequences within the NMB genome. The DNA sequences provide
CC further opportunities to find antigenic or immunogenic proteins which are
CC more effective in vaccines than the outer membrane proteins currently
CC used.
XX
SQ Sequence 349980 BP; 84410 A; 84863 C; 94187 G; 86520 T; 0 other;

```

Alignment Scores:

```

Pred. No.: 1.77e-13 Length: 349980
Score: 309.50 Matches: 119
Percent Similarity: 43.69% Conservative: 68
Best Local Similarity: 27.80% Mismatches: 186
Query Match: 15.33% Indels: 56
DB: 21 Gaps: 11
US-09-889-756a-2 (1-412) x AAF21607 (1-349980)

Qy 7 LysAlaMetArgAlaAlaAlaLeuAlaAlaValAlaLeuValLeuSerSerCysGly 26
Db 272470 AAAATGATGAATATGGCGGCTGTGCGCGGTCCGCGGCGCA----- 272511
Qy 27 LysGlyGlyAspAlaAlaGlnGlyGlyGlnProAlaGlyArgGlyAlaProAlaProVal 46
Db 272512 -----GGGGTTGGGGGGGATGCTTATCTGAAAGCCCGACGGCGTCT 272559
Qy 47 ValGlyValValThrValHisProGlnThrValAlaLeuThrValGluLeuProGlyArg 66
Db 272560 TATATTACGGAAACGGTCAGCGCGGCGACATCAGCCGAGCTTCTGCAACAGGGGAG 272619
Qy 67 LeuGluSerLeuArgThrAlaAspValArgAlaGlnValGlyIleIleGlnLysArg 86
Db 272820 ATTTCCGCCGTCACCTGCTATCGTCCGCGCGGCGGCGATCGGAGATTAAGATCTT 272679
Qy 87 LeuPheGlnGlnGlySerTyrValArgAlaGlyGlnProLeuTyrGlnIleAspSer--- 105
Db 272680 TATGTCAAACCTCGGGCAACAGGTTAAAGGCGCATTTGATGGCGAAATCATTCACACC 272739
Qy 106 -----SerThrTyrGlnLysLeuLeuGlnSerAlaArgAlaGlnLeu 119
Db 272740 TCCGACAGCAATACGCTCAATACGGAATAATCCAAAGTTGAGAAAGTATCAGCGAGCTG 272799
Qy 120 AlaThrAlaGlnAlaThrLeuAlaAlaAlaAspAlaAspLeuAlaArgTyrLysProLeu 139
Db 272800 GTGTCGACGAGTTCATTGTGGCGACGCGGAGAAAGAAATTAAGCGTCAGCGCGCTTA 272859
Qy 140 ValAlaAlaGlnAlaValaValaValaValaValaValaValaValaValaValaVala 159
Db 272860 TCGAAGAGAAACGCGACTTCGCAAGAGATTGGAAAGCGCGCAGATTCCTTGGCCGCC 272919
Qy 160 AlaGlnAlaGlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsn 179
Db 272920 GCCAAGCCATGTTGCGAGCTGACAGGCTTAAATCAGACAGCAAAATTCATCATCAT 272979
Qy 180 Arg-----SerArgIleThrAlaProIleSerGlyPheIleGly 192
Db 272980 ACCGCCGAGTCGAATTGGGCTACACGCGCATTCACCGACAGATGAGCGACGCTGGTG 273039
Qy 193 GlnSerLysValSerGlnGlyThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThr 212
Db 273040 GCGATTCCTCGTGAAGAGGGGCGACAGCTGTGAACCGCGG----- 273078
Qy 213 IleArgGlnThrAspProMetTyrValAsnValThrGlnSerAlaSerGluValMetLys 232
Db 273079 -----CAGTCTACGCGCGCATTTGTCATTCGATGGCAAT-----CTGGATATGATGTTG 273126
Qy 233 LeuArgArgGlnIleAlaGlnGlyLysLeu-----LeuAlaAlaAspGlyValIleAla 250
Db 273127 AACAAATGCGAGATTGCGAGGCGGATTTACCAAGAGGAGGCGGGCGAGATTTTTCG 273186
Qy 251 ValGlyIleLysPheAspAspGlyThrValTyrProGlnLysGlyArgLeuLeuPheAla 270
Db 273187 TTTACGATTTTGTCCGAACCGGATACG-----CCGATTAAGCGCAACCTGACACGCTC 273240
Qy 271 AspProValVal-----AsnGlnSerThrGlyGlnIleThr 282
Db 273241 GACCCCGGGCTGACACAGTGCCTGGCGGCTTACAAACACAGACAGATACGGCTTCC 273300
Qy 283 -----LeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetPro 296
Db 273301 AATCGGTCTACTATATATGATGATGCTGTTGTGTCGATTCGCGACGCGCAACCTGCGCACG 273360

```

```

Qy 297 GlyLeuTyrValArgValLeuMetAspGlnValAlaValaAspAsnAlaPheValPro 316
Db 273361 GGGATGACGACGAGAAATACGTTGAATGACGCGCTGAAATATGCTGATATATCCG 273420
Qy 317 GlnGlnAlaVal---ThrArgGlyAlaLysAspThrValMetIleValaAlaGlnGly 335
Db 273421 TCGCTGACCGTGAATAATCCGCGCGGCGAGCGCTTGTGTGCGGTGGTGGCGACGCG 273480
Qy 336 GlyMetGluProArgGluValThrValAlaGlnGlnGlnIleThrAspThrPheValThr 355
Db 273481 AAGCGCGGAGACCGAAATCCGACCGGATATGACACAGTATGATACCGAGTAA 273540
Qy 356 SerGlyLeuLysAspGlyAspLysValaValaGlnGlyIleSerIleAlaGlyIleThr 375
Db 273541 AGCGGCTTGAAGAAGGGGGGCAAAAGTGCATCTCCGAATTAACCGCGCGCA-GCAACA 273599
Qy 376 GlyAlaLysLysValThrProLysGluTyrPalaSerSerGluLeuGlnAlaAla--- 394
Db 273600 GGAAGCGGGAACGCGCCTTAGCGCGCGCGCGCGCATTAACGAATATGCGCTCTGA 273659
Qy 395 ProGlnSerGlyValGlnThrAla 402
Db 273660 ACACGGAACGTTTCAGACGCA 273683

RESULT 22
AA81490
ID AA81490 standard; DNA; 1437668 BP.
XX
AC AA81490;
XX
DT 04-DEC-2000 (first entry)
XX
DE N. meningitidis B full length genome DNA sequence SEQ ID NO:1068.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
KM antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW Meningococcus B; MenB; ds.
XX
OS Neisseria meningitidis.
XX
PN W0200022430-A2.
XX
PN 20-APR-2000.
XX
PF 08-OCT-1999; 99WO-US23573.
XX
PR 09-OCT-1998; 98US-0103794.
XX
PR 30-APR-1999; 99US-0132068.
XX
(CHIR ) CHIRON CORP.
XX
PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
PI Maignani V, Galeotti C, Mora M, Ratti G, Scarfelli M, Scariato V;
PI Rappuoli R, Pizza M;
XX
DR WPI; 2000-318079/27.
XX
PT Isolated nucleotide sequences of Neisseria meningitidis which can be
PT used in the diagnosis and treatment of N. meningitidis infection and
PT other Neisserial infections, for example, N.gonorrhoea -
XX
PS Claim 7; Page 866-1272; 1760pp; English.
XX
CC The present invention describes methods of obtaining immunogenic
CC proteins from Neisseria genomic sequences. AA81453 to AA82414
CC represent specifically claimed Neisseria meningitidis genome DNA
CC sequences; AA81260 to AA81303 and AA825620 to AA825663 represent
CC Neisseria DNA sequences and their corresponding proteins; AA81254 to
CC AA81259 and AA81304 to AA81321 represent PCR primers used in the
CC isolation of Neisseria meningitidis DNA sequences; and AA81322 to
CC AA81452 represent Neisseria meningitidis MenB polynucleotide ORF
CC sequences, which are all used in the exemplification of the present
CC invention. The nucleic acid sequences, protein sequences, and antibodies

```

CC against them, can be used in the manufacture of a composition. The
 CC composition can be used as a medicament (or in the manufacture of a
 CC medicament) for treating, preventing or diagnosing infection due to
 CC Neisserial bacteria. For example, some of the identified proteins could
 CC be components of vaccines against Meningococcus B: against all serotypes;
 CC and/or against all pathogenic Neisseriae. Identification of sequences
 CC from the bacterium will also facilitate production of biological probes,
 CC particularly organism-specific probes. Attempts to make efficacious
 CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
 CC Multivalent vaccines have also been tried but none have successfully
 CC overcome antigenic variability. The provision of further, complete
 CC sequences may provide an opportunity to identify secreted or surface
 CC exposed proteins that may be presumed targets for the immune system and
 CC which are not antigenically variable or at least more conserved than
 CC other more variable regions.

XX Sequence 1437668 BP: 344338 A; 353206 C; 365074 G; 355045 T; 5 other;

Alignment Scores:

Pred. No.:	9.81e-13	Length:	1437668
Score:	309.50	Matches:	119
Percent Similarity:	43.69%	Conservative:	68
Best Local Similarity:	27.80%	Mismatches:	186
Query Match:	15.33%	Indels:	56
DB:	21	Gaps:	11

US-09-889-756a-2 (1-412) x AA81490 (1-1437668)

QY 7 LysAlaMetArgAlaAlaLeuAlaAlaValAlaLeuValLeuSerSerCysGly 26
 Db 572470 AAAATGATGAATGATGGCGCGGTTCGCGCGCGCGCA----- 572511
 QY 27 LysGlyLeuAspAlaAlaGlnGlyGlyGlnProAlaGlyArgGluAlaProAlaProVal 46
 Db 572512 -----GCGGTTGGGCGCGCGATGCTTATCTGAAGCCCGAGCCGAGCGTCT 572559
 QY 47 ValGlyValValThrValHisProGlnThrValAlaLeuThrValGluLeuProGlyArg 66
 Db 572560 TATATTACGGAACGCTCAGCGCGCGCATCATCGCGGCTTCTGCAACAGGGAG 572619
 QY 67 LeuGlnSerLeuArgThrAlaAspValArgAlaGlnValGlyGlyLeuGlnLysArg 86
 Db 572620 ATTTCCGCGTCCAACTCGTATCGTGGCGCGCGCATCGCGGAGATTAAATATCTT 572679
 QY 87 LeuPheGlnGlnGlySerTyValArgAlaGlyGlnProLeuTyGlnIleAspSer--- 105
 Db 572680 TATCTCAACTCGCGGCAACAGTTAAAGGCGATTGATGGCGAAATCAATTCCACC 572739
 QY 106 -----SerThrTyGlnAlaAsnLeuGlnSerAlaArgAlaGlnLeu 119
 Db 572740 TCGCAGACCAATACGCTCATACGGAATAATCCAAAGTTGAAACGATCAAGCGAACCTG 572799
 QY 120 AlaThrAlaGlnAlaThrLeuAlaValAlaAspAlaAspLeuAlaArgTyLysProLeu 139
 Db 572800 GTGTCGCGACAGATTGATGGCGCGCGAGAGAAATATACCTTCAGCGCGCTTA 572859
 QY 140 ValAlaAlaGlnAlaValSerArgGlnGluTyArgAlaAlaValThrAlaLysArgSer 159
 Db 572860 TCGAAGAAAACCGCACTTCCAAAGAGATTGGAAAAGCGGACGAGATGGCTTGGCGCC 572919
 QY 160 AlaGlnAlaGlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsn 179
 Db 572920 GCCAAAGCCATGTTGCCGAGCTTATACGACAGCAAAATTTCCATCAAT 572979
 QY 180 Arg-----SerArgIleThrAlaProIleSerGlyPheIleGly 192
 Db 572980 ACCGCCAGTTCGGAATTGGGCTACACCGCATTAACCCCAACGATGCGCGACGCTGGTG 573039
 QY 193 GlnSerLysValSerGlyLysThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThr 212
 Db 573040 GCGATTCTCGTGAAGAGGGGACACTGTGAACCGCGC----- 573078
 QY 213 IleArgGlnThrAsnProMetTyValAsnValThrGlnSerAlaSerGluValMetLys 232

Db 573079 -----CAGTCTACCGCGAGATTGCTCAATTGGCGAAT-----CTGATATGATGTG 573126
 QY 233 LeuArgArgGlnIleAlaGlnGlyLysLeu-----LeuAlaAlaAspGlyValIleAla 250
 Db 573127 AACAAATGCAAGATGCGAGGGCGGATTTACCAAGGTGAAGGGCGGCGAGATATTTCG 573186
 QY 251 ValGlyIleLysPheAspAspGlyThrValTyProGluLysGlyArgLeuPheAla 270
 Db 573187 TTTACGATTTTGTGTCGCAACCGGATACG-----CCGATTAGCGGAAGCTGACAGGCTC 573240
 QY 271 AspProValVal-----AsnGlnSerThrGlyGlnIleThr 282
 Db 573241 GACCCCGCGTGCACACAGATGCTCGCGCGGTACCAACAGCAGTACGATACGCTTCC 573300
 QY 283 -----LeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetPro 296
 Db 573301 AATGCGGTCTACTATTATATGCGCTTGTGTGCGCAATCCGACGGCAAACTGCGCACG 573360
 QY 297 GlyLeuTyValArgValLeuMetAspGlnValAlaValAspAsnAlaPheValPro 316
 Db 573361 GCGATGACGACGACGAATACGTTGAATCGACGCGTGAATAATGCTGATATTCCG 573420
 QY 317 GlnGlnAlaVal---ThrArgGlyAlaLysAspThrValMetIleValAsnAlaGlnGly 335
 Db 573421 TCGCTGACCGTGAATAATCGCGCGCGCAAGCGCTTGTGCGCGTGGTGGCGACGCG 573480
 QY 336 GlyMetGlnProArgGluValThrValAlaGlnGlnGlnIleThrAsnThrPileValThr 355
 Db 573481 AAGCGCGCGGAACGCAAAATCCGACCGGTATGAGACAGTATGATGCAAGTAAAA 573540
 QY 356 SerGlyLeuLysAspGlyAspLysValValAlaGlnGlyIleSerIleAlaGlyIleThr 375
 Db 573541 AGCGGTTGAAGAAGGGGACCAAGTGTATCTCCGAATTAACCGCGCGCA-GCAACA 573599
 QY 376 GlyAlaLysLysValThrProLysGluTyPheLysSerSerGluAsnGlnAlaAla--- 394
 Db 573600 GGAAGCGCGCAACGCGCCTTAGCGCGCGCGCGCGCGATTAACGAATATGCGCTGA 573659
 QY 395 ProGlnSerGlyValGlnThrAla 402
 Db 573660 ACAAGGAACGTTTCAGACGCA 573683

RESULT 23
 AAS97298
 ID AAS97298 standard; DNA; 1176 BP.
 AC AAS97298;
 XX 12-MAR-2002 (first entry)
 DE Neisseria meningitidis virulence gene #103.
 XX
 KM Meningitis; virulence; gene; antibacterial; vaccine; veterinary;
 XX infection; Gram-negative bacteria; antimicrobial; ds.
 OS Neisseria meningitidis.
 XX
 PN WO200185772-A2.
 PD 15-NOV-2001.
 XX
 PF 08-MAY-2001; 2001MO-GB02003.
 PR 08-MAY-2000; 2000GB-0011108.
 PA (MICR-) MICROSCIENCE LTD.
 XX
 PI Tang C;
 XX
 DR WPI; 2002-066593/09.
 DR P-PSDB; AA073013.
 XX

PT meningitis and for identifying antimicrobial drug -
XX
PS Claim 1; Page 224-226; 423pp; English.
...

XX The invention relates to a peptide (I)
CC

Neisser meningitidis including virulence genes, or a related molecule having a 40% sequence similarity at the peptide or nucleotide level in Gram-negative bacterium, or its functional fragment, for therapeutic or diagnostic use. (I) and (II) are useful in the manufacture of a medicament for treating or preventing a condition (e.g., meningitis) associated with infection by *Neisseria* or Gram-negative bacteria. The product is useful for veterinary treatment and in a screening assay for the identification of an antimicrobial drug. The vaccines have prophylactic applications. AAS971196-AAS97305 represent *N. meningitidis* virulence genes and related PCR primers of the invention.

Sequence 1177 BP; 304 A; 276 C; 379 G; 218 T; 0 other;

Alignment Scores:

Pred. No.:	2,39e-16
Score:	308.00
Percent Similarity:	43.57%
Best Local Similarity:	27.14%
Query Match:	15.26%
DB:	24
	Gaps:
	11

US-09-889-756A-2 (1-412) X AAS97248 (1-1177)

Qy	1	LySLAmeWArGyLaLaLaLeuLaLaLaLaLaLaLeuValLeuSerSerCysGly	26
Db	7	AAATATGATGAATGGGCGGCTGTGGCGGCGCGCGCGCA-----	48
Qy	27	LyEGlyGLyAspAlaalaGnglyGlyGlnProAlaagLYaRGyuaLaproAlaproVal	46
Db	49	-----GCGGTGTTGGGCGGATGCTTATCTGAAGCCGAGCCGAGCTGCT	96
Qy	47	ValGLyValValThrValHisProGlnThrValAlaLeuThrValGluLeuProGLYArg	66
Db	97	TATATTACGGAACGGTCAAGCGCGGCGGCATCAAGCGGACGGTTTCTGCAACAGGGAG	156
Qy	67	LeuGLuSerLeuArgThrAlaAspValaArgaGlnValGLyGlyLeuLeuGlnLYaArg	86
Db	157	ATTTCGCCGTCCAACTGGATTGCGTCGCGCGCGCAGCATCGGGCAGATTAAAGAACTT	216
Qy	87	LeuPheGlnGluGlySerTyValaArgaGLyGlnProLeuTyRGlnIleAspSer---	105
Db	217	TATGTCAAACTCGGCGACAGGTTAAAAAGGCGCATTTGATTGATCGCGAATAATTCAACC	276
Qy	106	-----SerThyTyRGlnAlaLeuLeuGluSerLaArgAlaGlnLeu	119
Db	277	TGCGACAGCCAAATACGTCATACAGAAAAATCCAAATTGGAAACGTATCAGGCGAAGCTG	336
Qy	120	AlaThrAlaGlnAlaThrLeuAlaLYaSaLaAspLaLaAspLeuAlaArgTyLYaProLeu	139
Db	337	GTGTCGCGCACAGATTGCTATTGGGCGACCGCGGAGAAATATTAAGCTCAAGCGCGCTTG	396
Qy	140	ValAlaLaGlnValaValSerArgGlnGluTyraAspAlaAlaValaThrAlaLYaArgSer	155
Db	397	TGGAAGATGATGCGACCGCTTAAGAAGATTGGAAAGCGCACAGATGCGCTTGGCCGC	456
Qy	160	AlaGlnValaGLyValLYaSaLaalaGlnAlaalaIleLYaSerLaagLYaIleAsnLeuAsn	179
Db	457	GCCAAAGCCAAATGTTCCGAGCTGAAGGCTTAATCAACAGAGCAAAATTTCCATCAAT	516
Qy	180	Arg-----SerArgIleThrAlaProIleSerGlyPheIleGly	192
Db	517	ACCGCCGATCGGAATTGGGCTTACACGGCGCATTCAGCAACATGACAGCGACGCGTGGT	576
Qy	193	GlnSerLYaSaValSerGlnLYaThrLeuLeuAsnAlaGLyAspThrThrValLeuAlaThr	212
Db	577	GCGATTCTCGTGAAGAAGGGCGACACTGAAACGCGGG-----	615
Qy	213	IleArgGlnThrAsnProMetCtyValaAsnValThrGlnSerAlaSerGluValMetLYs	232

Db	616	-----CAGTGTACGCCGACGATGTCATTTGCCGAAT-----CTCGATATATATGTC	663
Qy	233	LeuArgrgInIlealaglucgLyLeu-----LeuIlaIaIaepglValIleala	250
Db	664	AAcAAATGcAGATGCGAGGGCGGATATTACcAAAGTGAAGGGCGAGGATATTTCG	723
Qy	251	ValglYIlelypheaapspglYThValYTrProgluysglYagLeuIleuheaIa	270
Db	724	TTTACGATTTTGTTCGAAcCGGATACG-----CCGATTAAggGAAgCTTCAGACGGCTC	777
Qy	271	AspProValVal-----AsnIuserPhrgIInIleThr	282
Db	778	GACCCCGGCGTACcCAGATGTCGTGGGGCGCTACAAcGAcGTACGATACGGCTTCC	837
Qy	283	-----LeuArglalaIaValProAsnAspGlInasnlleLeuMetPro	296
Db	838	AAATCGGCTACTATTATTCGCCGTCGTTGTGCGGAATCCGACGGCGAAATCGCCACG	897
Qy	297	GlYleuTrYValaIaIaValleuMetAspGlInValaIaValaspaanaIaheValValPro	316
Db	898	GGGATGAcGAcCGcCAGATATACGTTGAATGCAGCGGTGTGAaAATGTGCTATTATTCG	957
Qy	317	GlInIInIaIaIaIa-----ThrArgglYalIaLysasphrValmetIleValasnaIaIInglY	335
Db	958	TCGCTGAcCGTGAAaAATCCGGCGGcGAcGGCGCTTGTGCGCGTGGGGTCAGACGGC	1017
Qy	336	GlYMetGlUpRoArggluValThrValaIaIaIInglInglYThrAsnTrpIleValThr	355
Db	1018	AAcGGCGCGGAcCGcAAATTCcGAcCGGATAGAcAGAcAGTAcGATATCCGAAGTAAa	1077
Qy	356	SerGlYleuLysaspglYasplYasplYasplYValValcIgluYIleSerIlealaglyIleThr	375
Db	1078	AGCGGATTGAAGAGGGGGGAcCAAGTGTGTCATCTCCAAATTAACCGCGCGGAc-GcAGCA	1136
Qy	376	GlYalalYsIaYsValThrProLysgluTrpIaIaSerSergluSngIInIaIaIaIaPro	395
Db	1137	GGAAAGCGGcGAcACCGCCCTTAcG-----CGGCGCGCGCGCGC	1175
RESULT 25			
AA12305			
ID	AA12305	standard; DNA; 1179 BP.	
XX	AA12305;		
XX	08-OCT-1999	(first entry)	
XX	Neisseria meningitidis strain A complete ORF5 sequence.		
XX	Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;		
XX	treatment; Neisseria infection; meningitis; septicæmia; gonorrhea; ss.		
XX	Neisseria meningitidis.		
XX	W09924578-A2.		
XX	20-MAY-1999.		
XX	09-OCT-1998;	98W0-IB01665.	
XX	01-SEP-1998;	98GB-0019016.	
XX	06-NOV-1997;	97GB-0023516.	
XX	14-NOV-1997;	97GB-0024190.	
XX	18-NOV-1997;	97GB-0024386.	
XX	27-NOV-1997;	97GB-0025158.	
XX	10-DEC-1997;	97GB-0026147.	
XX	14-JAN-1998;	98GB-0000759.	
XX	(CHIR-) CHIRON SPA.		
XX	Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;		
XX	WPI; 1999-327407/27.		

DR P-PSDB; AAY38883.
 XX Proteins from *Neisseria meningitidis* and *N. gonorrhoeae* useful for
 PT diagnosis, treatment and prevention of infection
 XX
 PS Claim 9, Page 428, 524pp; English.
 XX
 CC Nucleotide sequences AA211972-212358 represent open reading frames
 CC (ORFs) of *Neisseria meningitidis* and *N. gonorrhoeae* which encode
 CC antigenic proteins (see AAY38849-438944). The antigenic proteins, their
 CC fragments, their nucleic acids and antibodies are used for diagnosis,
 CC prevention (as vaccines) or treatment of *Neisseria* infections,
 CC such as meningitis, septicemia and gonorrhea. Both organisms
 CC are closely related. Fragments of the nucleic acids are useful
 CC as hybridisation probes and antisense reagents.
 XX
 SQ Sequence 1179 BP, 306 A, 276 C, 379 G, 218 T, 0 other;
 Alignment Scores:
 Pred. No.: 2,39e-16 Length: 1179
 Score: 308.00 Matches: 114
 Percent Similarity: 43.57% Conservative: 69
 Best Local Similarity: 27.14% Mismatches: 176
 Query Match: 15.26% Indels: 62
 Gaps: 11
 US-09-889-756a-2 (1-412) x AA212305 (1-1179)
 QY 7 LysAlaMetArgAlaAlaLeuAlaAlaValAlaLeuValLeuSerSerCysGly 26
 Db 7 AAAATGATGAATGATGCGCTGTTGCGCGCGCGCGCA----- 48
 QY 27 LysGlyAlaPAPAlaAlaGlnGlyGlnProAlaGlyArgGlnAlaProAlaProVal 46
 Db 49 -----GCGTTGGGCGGATGCTTATCTGAAGCCCGAGCCGAGCTGCT 96
 QY 47 ValGlyValValThrValHisProGlnThrValAlaLeuThrValGlnLeuProGlyArg 66
 Db 97 TATATTACGAAACGGTCAGGCGCGCGACATCATCCGCGATTCTTCAACAGCGGAG 156
 QY 67 LeuGlnSerLeuArgThrAlaAspValArgAlaGlnValGlyGlyLeuGlnLysArg 86
 Db 157 ATTTCGCGCTCAACCTGATCGTCGCGCGCGCGATCGGCGATTAAGAACTT 216
 QY 87 LeuPheGlnGlnGlySerTyrValArgAlaGlnProLeuTyrGlnIleAspSer--- 105
 Db 217 TATGTCAAACTCGGCAACAGGTAAAGGCGCATTTGATTCGGAATCATTCGACC 276
 QY 106 -----SerThrTyrGlnAlaAsnLeuGlnSerAlaArgAlaGlnLeu 119
 Db 277 TCGCAGACCAATACGCTCATATACGAAATCCAAATTGGAACGATATCAGCGAGCTG 336
 QY 120 AlaThrAlaGlnAlaThrLeuAlaLysAlaAspLeuAlaArgTyrLysProLeu 139
 Db 337 GTGTGCGACAGATGTGATTGGGCGCGCGGAAAGAAATTAAGCTTCAGCGCGTTG 396
 QY 140 ValAlaAlaGlnAlaValSerArgGlnLysThrAspAlaAlaValThrAlaLysArgSer 159
 Db 397 TCGAAGGATGATGACCGCTTAAGAAAGATTGGAACGCGACAGATGCTGCGCCG 456
 QY 160 AlaGlnAlaGlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyTyrAsnLeuAsn 179
 Db 457 GCCAAAGCATTTGTCGCGAGCTCATATCAGACAGCAAAATTTCCATCAT 516
 QY 180 Arg-----SerArgIleThrAlaProIleSerGlyPheIleGly 192
 Db 517 ACCCGCGATGCGAATTGGCTACCGCGCATTCACGCAAGATGAGCGACGATGATG 576
 QY 193 GlnSerLysValSerGlnGlyThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThr 212
 Db 577 GCGATTCTCTGGAAGAGGCGCAACTGTGAACGCGCG----- 615
 QY 213 IleArgGlnThrAsnProMetTyrValAsnValThrGlnSerAlaSerGlnValMetLys 232

Db 616 -----CAGTCTACCGCAGCATTTGCTCAATTGGCGAAT-----CTGATATGATGTTG 663
 QY 233 LeuArgArgGlnIleAlaGlnGlyLysLeu-----LeuAlaAlaAspGlyValIleAla 250
 Db 664 AACAAATGCAATGTCGCGAGGATATTAACAAAGTGAAGCGGCGCAGATATTTCG 723
 QY 251 ValGlyIleLysPheAspAspGlyThrValTyrProGlnLysGlyArgLeuPheAla 270
 Db 724 TTTACGATTTTGTTCGCAACCGGATACG-----CCGATTAAAGCGAAGCTGCACAGGTC 777
 QY 271 AspProValVal-----AsnGlnSerThrGlyGlnIleThr 282
 Db 778 GACCCCGCGTACACAGATGCTCGCGCGCTACAAACAGCATGATACCGCTTCC 837
 QY 283 -----LeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetPro 296
 Db 838 AATGCGGTCTACTATTATATGCCCCGTTGTTGTGCGGAATCCGACGCGCAAACTCGCACG 897
 QY 297 GlyLeuTyrValArgValLeuMetAspGlnValAlaValAspAsnAlaPheValPro 316
 Db 898 GCGATGACGACGACGATACGTTGAATGACCGGTGTGAATAATGCTGATATTTCG 957
 QY 317 GlnGlnAlaVal--ThrArgGlyAlaLysAspThrValMetIleValAsnAlaGlnGly 335
 Db 958 TCGCTGACCGTGAATAATCGCGCGCGCGCGCTTGTGCGCGTGTGGGTGACAGCGCG 1017
 QY 336 GlyMetGluProArgGlnValThrValAlaGlnGlnGlnLysThrAsnThrPileValThr 355
 Db 1018 AAGCGCGCGGAACCGCAAAATCCGACCGGTATGAGACAGTATGATACCGAAGTAAA 1077
 QY 356 SerGlyLeuLysAspGlyAspLysValValIleGlnGlyIleSerIleAlaGlyIleThr 375
 Db 1078 AGCGGTGTAAGAGGGGGAAGAGTGTATCTCCAAATTAACCGCCCGCA-GCAGCA 1136
 QY 376 GlyAlaLysLysValThrProLysGluThrPalaSerSerGlnAsnGlnAlaAlaPro 395
 Db 1137 GAAAGCGCGGAACGCGCCTAG-----CGCCCGCGCGCGCG 1175
 RESULT 26
 AA253631 standard; DNA, 1179 BP.
 ID AA253631
 AC AA253631;
 XX
 DT 21-MAR-2000 (first entry)
 XX
 DE *Neisseria gonorrhoeae* ORF 290 partial DNA sequence SEQ ID NO:1211.
 XX
 KW *Neisseria meningitidis*; *Neisseria gonorrhoeae*; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KW antibacterial; gene therapy; ds.
 OS *Neisseria gonorrhoeae*.
 PN WO9957280-A2.
 PD 11-NOV-1999.
 XX
 PF 30-APR-1999; 99WO-US09346.
 XX
 PR 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX

PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizzia M, Rappuoli R, Ratcl G, Scalato E, Scarselli M;
 PI Tectelin H, Venter JC;
 XX
 XX WPI; 2000-062150/05.
 DR P-PSDBJ; AAY74871.
 XX
 PT Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -
 XX
 XX Claim 7; Page 678; 1453pp; English.
 XX
 CC AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of.
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenetic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 XX
 SO Sequence 1179 BP; 306 A; 276 C; 379 G; 218 T; 0 other;
 Alignment Scores:
 Pred. No.: 2,39e-16 Length: 1179
 Score: 308.00 Matches: 114
 Percent Similarity: 43.57% Conservative: 69
 Best Local Similarity: 27.14% Mismatches: 176
 Query Match: 15.26% Indels: 62
 DB: 21 Gaps: 11
 US-09-889-756a-2 (1-412) x AA253633 (1-1179)
 QY 7 LysAlaMetArgAlaAlaAlaLeuAlaAlaAlaValAlaLeuValLeuSerSerCysGly 26
 Db 7 AAAATGATGAATGGCGCGCTGTTGCGCGCGCGCGCA----- 48
 QY 27 LysGlyGlyAspAlaAlaGlnGlyGlyProAlaGlyArgGlyAlaProAlaProVal 46
 Db 49 -----GGCGTTGGGGCGGATGCTTATTCGAAGCCGACCCGACGGCTGCT 96
 QY 47 ValGlyValAlaThrValHisProGlnThrValAlaLeuThrValGluLeuProGlyArg 66
 Db 97 TATATTACGAAGACGGTCAGCGCGCGACATCAGCGGACGGTTTCGCAACAGGGAG 156
 QY 67 LeuGluSerLeuArgThrAlaAspValArgAlaGlnValGlyGlyLeuGlnLysArg 86
 Db 157 ATTTCCGCGCTCAACCTGTGATCGGTCCGCGCGACGATCGGGGCAATTAAGAACTT 216
 QY 87 LeuPheGlnGluGlySerTyrValArgAlaGlyGlnProLeuTyrGlnHisLeuAspSer--- 105
 Db 217 TATGTCAAACTCGGCAACAGGTTAAAGGCGCATTTGATTGCGGAATCAATTTCAGACC 276
 QY 106 -----SerThrTyrGluAlaLeuLeuGluSerAlaArgAlaGlnLeu 119
 Db 277 TCGCAGACCAATACGCTCAATACGAAATTCCTCAATTGGAAAGCTATACAGCGAAGCTG 336
 QY 120 AlaThrAlaGlnAlaThrLeuAlaLysAlaAspAlaAspLeuAlaArgTyrLysProLeu 139

Db 337 GTGTCCGACAGATTCGATTCGGCAGCGAGAGAAATATACCTTCAGCCGCGTGG 396
 QY 140 ValAlaAlaGluAlaValSerArgGlnGlyTyrAspAlaAlaValThrAlaLysArgSer 159
 Db 397 TGGAAAGATGATGGACCGCTTAAAGAAATTTGGAAAGCGCACGATTCGCTTGGCCGC 456
 QY 160 AlaGluAlaGlyValLysAlaAlaGlnAlaAlaLeuSerAlaGlyLeuAsnLeuAsn 179
 Db 457 GCCAAGCAATGTTGCCAGCTGACGCTCTATCAGACGACCAAAATTCATCCATCAT 516
 QY 180 Arg-----SerArgIleThrAlaProIleSerGlyPheIleGly 192
 Db 517 ACCCGCAGTCGAATTGGGCTACACCGCATTAACCAAGATGACGACGACGCGGTGG 576
 QY 193 GlnSerLysValSerGluGlyThrLeuLeuAsnAlaGlyAspThrValLeuAlaThr 212
 Db 577 GCGATTCTCGTGAAGAGGGCAGACTGTGAACCGCGC----- 615
 QY 213 IleArgGlnThrAspProMetTyrValAsnValThrGlnSerAlaSerGlyValMetLys 232
 Db 616 -----CAGTCTACGCCGACGATTCGCAATTCGCAAT-----CTGATATGATGTTG 663
 QY 233 LeuArgArgGlnIleAlaGlyLysLeu-----LeuAlaAlaAspGlyValIleAla 250
 Db 664 AACCAAAATGACAGATTGCCGAGGCGGATTAACCAAGGTGAAGCGGCGCAGATATTGC 723
 QY 251 ValGlyLysPheAspAspGlyThrValTyrProGluLysGlyArgLeuLeuPheAla 270
 Db 724 TTTTCGATTTCCTCCGAACCGGATAC-----CCGATTAAAGGGAAGCTGCACAGCTC 777
 QY 271 AspProValVal-----AsnGluSerThrGlyGlnIleThr 282
 Db 778 GACCCCGGCGTACACAGATGTGTCGGCGGTACAAAGCAGTACGATACGATTCGCTCC 837
 QY 283 -----LeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetPro 296
 Db 838 AATCGGCTACTATTATATGCGCCGTTGCTGTCGGAATCCGACGCAAACTTCGCCACG 897
 QY 297 GlyLeuTyrValArgValLeuMetAspGlnValAlaValAspAsnAlaPheValValPro 316
 Db 898 GGGATGACGACGCGGAATACGTTGAATACGACGCTGTGAATAATGCTGATTTATTCG 957
 QY 317 GlnGlnAlaVal--ThrArgGlyAlaLysAspThrValMetIleValAsnAlaGlnGly 335
 Db 958 TCGCTGACCGTGAATAATTCGCGCGCAGCGCGCTTGTGCGGCTTGGCGACAGCGGC 1017
 QY 336 GlyMetGluProArgGluValThrValAlaGlnGlnGlnGlyThrAsnTrpIleValThr 355
 Db 1018 AAGCGCGCGGAACGCGAAATCCGACCCGATGAGACAGATGATGATACGAAATGAAA 1077
 QY 356 SerGlyLeuLysAspGlyAspLysValValValGluGlyIleSerIleAlaGlyIleThr 375
 Db 1078 AGCGGGTTGAAGAGGGGACAAAGTGTCACTCCGAATAATACCGCGCGCA-GCAGCA 1136
 QY 376 GAlaLysLysValThrProLysGluTrpAlaSerSerGluAsnGlnAlaAlaPro 395
 Db 1137 GGAAGCGCGGAACGCCCTTAG-----CGGCCGCGCGCCCG 1175
 RESULT 28
 AAS74537/c
 ID AAS74537 standard; cDNA; 3732 BP.
 AC AAS74537;
 DT 13-FEB-2002 (first entry)
 DE DNA encoding novel human diagnostic protein #10341.
 XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; se.
 XX Homo sapiens.
 OS

PS Claim 1; SEQ ID No 28696; 103bp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations in
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3732 BP; 833 A; 1010 C; 961 G; 928 T; 0 other;
Alignment Scores:
Pred. No.: 1.88e-15 Length: 3732
Score: 304.50 Matches: 73
Percent Similarity: 51.22% Conservatave: 53
Best Local Similarity: 29.67% Mismatches: 83
Query Match: 15.08% Indels: 37
DB: 23 Gaps: 5
US-09-889-756a-2 (1-412) x AAS92892 (1-3732)
QY 184 ThrAlaProIleSerGlyPheIleGlyIleSerGlyValSerGlyValThrLeuLeuAsn 203
DB 2562 ACCTTCCCGGTCAACGGGATTCACCCGCAATCGAATCTTCAATGCTGGAC 2503
QY 204 AlaGlyAspThrValLeuAlaThrIleArgIleThrAsnProMetCysVal 221
DB 2502 -----CTTCGAAACCACTGGCGTATATGTCGGGAA 2470
QY 222 -----AsnValThrGlnSerIleSerGlyValMetLeuLeuArg 235
DB 2469 TCCATATGGAATCTCTGCTCACTACAGATCTTCCAGCTTCTGCGCTGAAACAG 2410
QY 236 GlnIleAlaGlyIleValLeuAlaIleAspGlyValIleAlaValGlyIleAspHe 255
DB 2409 GAACCTGGGATGGACGCTGAACAGAAAGCAAGCAAGTCTACTGATCAC 2350
QY 256 AspAspGlyThrValIleProGlyIleGlyValLeuLeuPheAlaAspProValValAsn 275
DB 2349 AGTACCGCATTAATGTTCCCGACGACGCTACGCTGAAATCTCTGACGTACCGTTGAT 2290
QY 276 GluSerThrGlyGlnIleThrLeuAlaIleAlaValProAsnAspGlnAsnIleLeuMet 295
DB 2289 CAGACCACTGGGTCTATACCTTACCGCTTCTCTCCGAAACCGGATCACACTCTGCTG 2230
QY 296 ProGlyLeuValValAlaValLeuMetAspGlnValAlaValAspAsnAlaPheValVal 315
DB 2229 CCGGGTATGTTCTGTCGCGACGCTGTGGAAGAAGGCTTATCCAAACGCTATTTTGTG 2170
QY 316 ProGlnIleAlaValThrArgIleAlaAsp-----ThrValMetIleValAsnAla 333
DB 2169 CCGGAACGCGGCGTAAACCGTACCGCGGTGCGCATGACCGCACTGCTGTTGGCGCG 2110
QY 334 GlnGlyIleMetGluProArgGluValThrValAlaGlnGlnGlnGlnIleThrAsnTyrIle 353
DB 2109 GATGACAAAGTGAACCCGCTCCGATCTTCCAAAGCCGCTATTTGGGATTAAGTGGCTG 2050
QY 354 ValThrSerGlyLeuIleAspGlyAspIleValValValGlnGlyIle----- 369

DB 2049 GTGACAGAAAGTCTGAAAGCAGCGATCGCGTAGTATAGTGGCGTCAAGAAAGTGGCT 1990
QY 370 -----SerIleAlaGlyIleThrGlyAlaIleValVal 380
DB 1989 CCGGTGTCCAGGGCCCGGCTTCCGACATACCCGACTTGACGGA----- 1942
QY 381 ThrProGlyIleThrAlaSerSerGlyValGlnAlaIleAlaIleProGlnSerGlyValGln 400
DB 1941 ---CCGAGCGAAATCGCGCGTGCAGAAAAGCACTGAGCGCGCCCGCACAGCAACAGGG 1885
QY 401 ThrAlaSerGlyAlaIleVal 406
DB 1884 GGGGACACACAAAGCCAAAG 1867
RESULT 30
AAS93885/C
ID AAS93885 standard; cDNA; 3732 BP.
AC AAS93885;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #29689.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001NO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217;
XX 23-AUG-2000; 2000US-0649167;
XX
XX (HSE-) HYSER INC.
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX P-PsDB; ABG29698.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -
PS Claim 1; SEQ ID No 29689; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations in
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 3732 BP; 833 A; 1010 C; 961 G; 928 T; 0 other;

Alignment Scores:

Pred. No.:	1.88e-15	Length:	3732
Score:	304.50	Matches:	73
Percent Similarity:	51.22%	Conservative:	53
Best Local Similarity:	29.67%	Mismatches:	83
Query Match:	15.08%	Indels:	37
DB:	23	Gaps:	5

US-09-889-756A-2 (1-412) x AAS93885 (1-3732)

```

Qy 184 ThrAlaProIleSerGlyPheIleGlyIleSerIleValSerGlyIleThrLeuLeuAsn 203
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2562 ACCCTCCCGGTCACAGCGGCAATTCACCCCAATGTAATCTTCAAAATGCTGGAC 2503
Qy 204 AlAGlyAspThrValLeuAlaThrIleArgIleThrAsnProMetIleVal----- 221
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2502 -----CTTTCGAAACACCTGGCCCTATATGTCGGGAAA 2470
Qy 222 -----AsnValThrGlnSerAlaSerGluValMetIleValArg 235
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2469 TCATATGCAATTCCTGGTCACTTACAGGATCTCCAAACGCTTCTCCGCTGAAACAG 2410
Qy 236 GlnIleAlaGluGlyIleLeuAlaAlaAspGlyValIleAlaValGlyIleValPhe 255
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2409 GAACCTGGCGAATGCGACCGTGAACAGAGAACGCGAAACCGAAGTGTACATGATCACC 2350
Qy 256 AspAspGlyThrValIleProGluValGlyIleValLeuPheValAsn 275
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2349 AGTACGCGCATTAAGTTCGCCAGAGACGCTACCGTGAATTTCTGACGTTACGTTGAT 2290
Qy 276 GluSerThrGlyIleThrIleValAlaValAlaValProAsnAspGlnAsnIleLeuMet 255
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2289 CACACCACTGGCTCTATACCTTACCGGCTATCTTCCGAAACCGGATCACACTGCTGCTG 2230
Qy 296 ProGlyLeuThrValArgValLeuMetAspGlyValAlaValAlaValAspAsnIlePheVal 315
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2229 CCGGGTATGTTCTGCGCGACGCTGGAAGAAAGGCTTATCAACACGCTATTTTATGTC 2170
Qy 316 ProGlnGlnAlaValThrArgGlyAlaValAsp-----ThrValMetIleValAsnAla 333
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2169 CCGCAACAGGCGCTAACCGCTACGCGCGCGGATGCCACCGACTCGATGATGGGCGC 2110
Qy 334 GlnGlyGlyMetGluProArgGluValThrValAlaGlnGlnGlnGlyThrAsnTrpIle 353
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2109 GATGACAAAGTGAACCCGCTCGAGTGTGCAAGCAGCGCTATTGCGCATTAAGTGGCTG 2050
Qy 354 ValThrSerGlyLeuIleValAspGlyValValValGluGlyIle----- 369
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2049 GTACACGAAGGTCTGAAGACAGCGGATCGGTATTAATGAGTGGCTGCAAAAGTGGCT 1990
Qy 370 -----SerIleAlaGlyIleThrGlyAlaValVal 380
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1989 CCGTGTGTCCAGGCGCCCGGCTTCCGCGACATACCGACCTAGACCGA----- 1942
Qy 381 ThrProIleGluTrpAlaSerSerGluAsnGlnAlaAlaProGlnSerGlyValGln 400
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1941 ---CCAGCGCAAAATCGCGCGCGCAAAAGCACTAGCGCGCCGCGCACGCAACCAAGCGG 1885
Qy 401 ThrAlaSerGluAlaVal 406
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1884 GGGGACACACACCAAGCCAG 1867

```

RESULT 31

AAA81495

ID AAA81495 standard; DNA; 44608 BP.

AC AAA81495;

DT 04-DEC-2000 (first entry)

XX N. meningitidis partial DNA sequence gnm_42 SEQ ID NO:42.

XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
 XX antigen; vaccine; diagnosis; infection; antibacterial; identification;
 KW Meningococcus B; MenB; ds.

XX Neisseria meningitidis.

XX WO200022430-A2.

XX 20-APR-2000.

XX 08-OCT-1999; 99WO-US23573.

XX 09-OCT-1998; 98US-0103794.

XX 30-APR-1999; 99US-0132068.

XX (CHIR) CHIRON CORP.

XX Frazier CM, Hickey E, Peterson J, Tettelin H, Venter JC;
 XX Maignant V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
 XX Rappelli R, Pizza M;

XX WPI; 2000-318079/27.

XX Isolated nucleotide sequences of Neisseria meningitidis which can be
 XX used in the diagnosis and treatment of N. meningitidis infection and
 XX other Neisseria infections, for example, N.gonorrhoea -
 XX Claim 7; Page 1283-1296; 1760pp; English.

XX The present invention describes methods of obtaining immunogenic
 XX proteins from Neisseria genomic sequences. AAA81453 to AAA82414
 XX represent specifically claimed Neisseria meningitidis genomic DNA
 XX sequences; AAA81260 to AAA81303 and AA825620 to AA825663 represent
 XX Neisseria DNA sequences and their corresponding proteins; AAA81254 to
 XX AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
 XX isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
 XX AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
 XX sequences, which are all used in the exemplification of the present
 XX invention. The nucleic acid sequences, protein sequences, and antibodies
 XX against them, can be used in the manufacture of a composition. The
 XX composition can be used as a medicament (or in the manufacture of a
 XX medicament) for treating, preventing or diagnosing infection due to
 XX Neisseria bacteria. For example, some of the identified proteins could
 XX be components of vaccines against Meningococcus B; against all serotypes;
 XX and/or against all pathogenic Neisseriae. Identification of sequences
 XX from the bacterium will also facilitate production of biological probes,
 XX particularly organism-specific probes. Attempts to make efficacious
 XX Meningococcus B vaccines have failed mainly due to antigen tolerance.
 XX Multivalent vaccines have also been tried but none have successfully
 XX overcome antigenic variability. The provision of further, complete
 XX sequences may provide an opportunity to identify secreted or surface
 XX exposed proteins that may be presumed targets for the immune system and
 XX which are not antigenically variable or at least more conserved than
 XX other more variable regions.

XX SQ Sequence 44608 BP; 10938 A; 10835 C; 11999 G; 10834 T; 2 other;

Alignment Scores:

Pred. No.:	3.79e-14	Length:	44608
Score:	304.50	Matches:	117
Percent Similarity:	44.34%	Conservative:	67
Best Local Similarity:	28.19%	Mismatches:	176
Query Match:	15.08%	Indels:	56
DB:	21	Gaps:	13

US-09-889-756A-2 (1-412) x AAA81495 (1-44608)

```

Qy 25 CysGlyIleValGlyIleValAlaGlnGlyIleValProAlaGlyArg----- 40
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 530 TCGCGC-----GATCGCGCGCGACGCGTGTGGCGCGGATGCTTATCT 574

```

```

Qy 41 GlnAlaProAlaProValValGlyValVal---ThrValHisProGlnThrValAlaLeu 59
Db 575 GAACCCGAGCGCGAGCTGCTTATATACGAAACGGTCAGCGCGGACATCATGACGCCG 634
Qy 60 ThrValGluLeuProGlyArgLeuGluSerLeuArgThrAlaAspValArgAlaGlnVal 79
Db 635 ACCGTTTCTCGACACGAGAGATTTCCGCTCAACCTGATGCGTCCGCGCCAGCA 694
Qy 80 GlyGlyIleIleGlnLysArgLeuPheGlnGluGlySerTyValArgAlaGlyGlnPro 99
Db 695 TCGGGGAGATTAAGATTAATCTTATGTCTAACTCGGGCAACGGTTAAAGGGCATTTG 754
Qy 100 LeuTyrglnIleAspSer-----SerThrTyrglnAlaAsnLeu 112
Db 755 ATTGGGAAATCAATTCGACCTCGACACCAATACCGTCATACGAAATAATCCAGTTG 814
Qy 113 GluSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaLysAlaAspAlaAsp 132
Db 815 GAAACGTATCAGCGGAGAGCTGCTGCGCACAGATTGCGACGCGCGAGAGCAA 874
Qy 133 LeuAlaArgTyrlvsProLeuValAlaAlaGlnAlaValSerArgGlnGluTyAspAla 152
Db 875 TATAAGCGTCAAGCGGCGCTTATGAGAGAAACCGGACTTCCAAAGAGATTGGAAAGC 934
Qy 153 AlaValThrAlaLysArgSerAlaGlnAlaGlyValLysAlaAlaGlnAlaAlaIleLys 172
Db 935 GCGCAGAGATGCGTTTCCGCCCGCAAAACCAATGTGCGCAGCTGAAGCTTATATCGA 994
Qy 173 SerAlaGlyIleAsnLeuAsnArg-----SerArgIleThrAla 185
Db 995 CAGAGCAAAATTTCCATCAATATACCGCCGATTCGATTCACGCGCATTAACGCA 1054
Qy 186 ProIleSerGlyPheIleGlyGlnSerLysValSerGlyGlnLeuLeuAlaGly 205
Db 1055 ACGATGACGCGCAGCGGTGCGCATTTCTGTGAAGGGGCGAGACTGTGAACGCGCGC 1114
Qy 206 AspThrThrValLeuAlaThrIleArgGlnThrAsnProMetTyValAsnValThrGln 225
Db 1115 -----CAGTCACGCGCAGATTTGCCATTTGGCGCAAT 1147
Qy 226 SerAlaSerGluValMetLysLeuArgArgGlnIleAlaGlnGlyLysLeu-----Leu 243
Db 1148 -----CTGGATATGATGTTGACAAACAAATGCAATGCGAGCGGCGATTAACCAAGGTG 1201
Qy 244 AlaAlaAspGlyValIleAlaValGlyIleLysPheAspArgLysThrValTyProGlu 263
Db 1202 AAGCGGGGCGAGATATTTCTTTACGATTTTGTCCGAACCGGATACG-----CCGATT 1255
Qy 264 LysGlyArgLeuLeuPheAlaAspProValVal-----Asn 275
Db 1256 AAGCGGAAGCTCGACAGCGTGCACCCCGGCTGCACACGATGCTGCGGCGGTTTACAC 1315
Qy 276 GluSerThrGlyGlnIleThr-----LeuArgAlaAlaValProAsn 289
Db 1316 AGCAGTACGAGTACGCGCTTCAATGCGGTCTACTATTATGCCCCGTTGTGCGCAT 1375
Qy 290 AspGlnAsnIleLeuMetProGlyLeuTyValArgValLeuMetAspGlnValAlaVal 309
Db 1376 CCGACCGGCAAACTCCGACCGGGGATGACGACGAGATACGATTGAATCGACGCGCTG 1435
Qy 310 AspAsnAlaPheValValProGlnAlaVal---ThrArgValAlaLysAspThrVal 328
Db 1436 AAAAATGCTGATTTATTCCTGCTGACCGTGAATAATCCGGGCGGCAAGGCGCTTTGTG 1495
Qy 329 MetIleValAsnAlaGlnGlyIleMetGluProArgGluValThrValAlaGlnGlnGln 348
Db 1496 CGCGTGTGGGTGGGAGACGCAAGCGCGGAGCGCAATCCGAGACCGATATAGAGAC 1555
Qy 349 GlyThrAsnTyrIleValThrSerGlyLeuLysAspGlyAspValValValGlnGly 368
Db 1556 AGTATGATATCCGAATAAAAGCGGTTGAAGAGGGGAGCAAAAGTGTCATCTCCGAA 1615

```

```

Qy 369 IleSerIleAlaGlyIleThrGlyAlaLysValThrProLysGluTyrPalasSerSer 388
Db 1616 ATAAACCGCCCGCA-GCAACGAGAAACCGCGAAGCGCCCTTAGCGCGCCGCGCCG 1674
Qy 389 GluAsnGlnAlaAlaAla---ProGlnSerGlyValGlnThrAla 402
Db 1675 ATAAACGAAATATGCGCTGACACGGAACGAGTTTTCAGACGCA 1719

RESULT 32
AAF28554
ID AAF28554 standard; DNA; 269223 BP.
XX
AC AAF28554;
XX
DT 04-APR-2001 (first entry)
XX
DE Genomic fragment #41.
XX
KW Genomic library; bacteria; human upper airway; otitis media; sinusitis;
KW bronchopulmonary; endocarditis; meningitis; ss.
XX
OS Moraxella catarrhalis.
XX
EN W0200078968-A2.
XX
PD 28-DEC-2000.
XX
PF 16-JUN-2000; 2000MO-US16649.
XX
PR 18-JUN-1999; 99US-0140121.
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Lagace RE, Patterson C, Berg KL;
XX
WP1; 2001-041427/05.
XX
PT Genomic library for identifying diagnostic and therapeutic
PT compositions, and for identifying virulence factors, regulatory
PT elements and drug targets, comprises Moraxella catarrhalis nucleic
PT acids -
XX
PS Claim 1; Page 486-545; 545pp; English.
XX
CC The present invention relates to a Moraxella catarrhalis genomic library
CC comprising of a combination of 41 nucleic acid molecules (see
CC AA28514-AA28554). The library has a number of uses described in the
CC specification e.g. is useful for identifying diagnostic and therapeutic
CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large
CC aerobic, gram-negative diplococcus, normally found among the bacterial
CC flora of human upper airways. M. catarrhalis is known to cause acute,
CC localised infections such as otitis media, sinusitis and bronchopulmonary
CC infection and life-threatening, systemic diseases including endocarditis
CC and meningitis.
XX
SQ Sequence 269223 BP; 77067 A; 56596 C; 57380 G; 78180 T; 0 other;
XX

Alignment Scores:
Pred. No.: 3.34e-13 Length: 269223
Score: 304.50 Matches: 85
Percent Similarity: 54.42% Conservative: 32
Best local Similarity: 39.53% Mismatches: 73
Query Match: 15.08% Indels: 25
DB: 22 Gaps: 6

US-09-889-756A-2 (1-412) x AAF28554 (1-269223)
Qy 21 ValLeuSerSerGlyGlyLysGly-----GlyAspAlaAlaGlnGlyGln 36
Db 238995 CTGCTCACCCCTTTGGCAACGAGATTAATACCTTGCGGTGCTACTACCGAGGTGTA 239054
Qy 37 ProAlaGlyArg-----GlnAlaPro----- 43

```


Db 23905 CTCACGGCGATCTATGGATGATCAACCAATCAAGACACCGCTCAAGACGCTCCCA 239114
Qy 44 -----AlapProValValGlyValValThyValHisProGlnThrValAlaLeuThr 60
Db 239115 AACCAAAACGACAGATGATGCGGTGGAAGCGGTACCTCCCAATCATGATCAATCAAA 239174
Qy 61 ValGlnLeuPro-----GlyArgLeuGlnSerLeuArgThyAlaAspValArgAlaGln 78
Db 239175 AATCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 239234
Qy 79 ValGlyValGlyLeu-----GlnGlnArgLeuPheGlnGlnGlnGlnGlnGlnGlnGln 97
Db 239235 ATCAGCGGTGTCGCGATGACGCGGTGTCGCGGTGTCGCGGTGTCGCGGTGTCGCGGT 239294
Qy 98 GlnProLeuValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 117
Db 239295 CAGGTGCTTGCACACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 239354
Qy 118 GlnLeuValThyAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 137
Db 239355 GAGCTTGCACACGCGGTGTCGCAAGAAAGCGGTGTCGCAAGAAAGCGGTGTCGCAAG 239414
Qy 138 ProLeuValAlaAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 157
Db 239415 CCATTGTTAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 239474
Qy 158 ArgSerAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 177
Db 239475 CCGGACAGTTGTTGGGAAATGTTACCGCTGCCAGCCCATGTCGAAAGTCTGTAAT 239534
Qy 178 LeuAsnArgSerArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 197
Db 239535 CAGTAGGAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 239594
Qy 198 GlnGlyThyLeuLeuAsnAlaGlyAspThyThyValLeuAlaThr 212
Db 239595 GTTGTGTCAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 239633
RESULT 33
ABQ24964/c
ID ABQ24964 standard; DNA, 607 BP.
XX AC ABQ24964;
XX 12-JUL-2002 (first entry)
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 11555.
XX
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KM drug; side effect; cancer; central nervous system; cardiovascular;
KM gastrointestinal; respiratory system; single nucleotide polymorphism;
KM SNP; cell differentiation; ds.
XX Homo sapiens.
XX WO200218632-A2.
XX 07-MAR-2002.
XX 01-SEP-2001; 2001WO-EP10074.
XX 01-SEP-2000; 2000DE-1043826.
XX 05-SEP-2000; 2000DE-1044543.
XX (EPIG-) EPIGENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K, Guefig D;
XX WPI; 2002-371829/40.
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA

XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridized to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridization to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridized to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's) and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 607 BP; 121 A; 53 C; 159 G; 274 T; 0 other;
XX
Alignment Scores:
Pred. No.: 2,24e-15 Length: 607
Score: 292.00 Matches: 60
Percent Similarity: 56.00% Conservative: 52
Best Local Similarity: 30.00% Mismatches: 78
Query Match: 14,46% Indels: 10
DB: 24 Gaps: 2
US-09-889-756a-2 (1-412) x ABQ24964 (1-607)
Qy 185 AlapProLeuSerGlyPheLeuGlnSerGlyValSerGlnGlyThyLeuAsnAla 204
Db 607 TCTCCGATTAAGATGCGCATTAATTAATGCAAGTAACGAAACGATTATTAACAAAC 548
Qy 205 GlyAspThrThyValLeuAlaThrLeuGlnThrAsnProMetGlyValAsnValThr 224
Db 547 GATCAACGACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 488
Qy 225 GlnSerAlaSerGlnValMetGlyLeuArgGlnLeuAlaGlnGlyLeuLeuAla 244
Db 487 CAATCCAAACGACCTCTCAAGCTTAACCAAACTTAACGAATTAACGCTTAACAA 428
Qy 245 AlapGlyValLeuAlaGlyLeuPheAspAspGlyThyValTyProGlnGly 264
Db 427 AAAAAGCAAAACCAAAATATGATCAATCAACCAATTAACGATTAATTCGCAAAAC 368
Qy 265 GlyArgLeuLeuPheAlaAspProValValAsnGlnSerThyGlnGlnLeuArg 284
Db 367 GATACCTTAATATCTTAACGTTACCTTAACCAACCACTTAATCTATCACCTACGC 308
Qy 285 AlAlaValProAsnAspGlnAsnLeuMetProGlyLeuValValArgValLeuMet 304
Db 307 GCTATCTTCCGACCGAATCACTCACTCACTCACTCACTCACTCACTCACTCACTCA 248
Qy 305 AspGlnValAlaValAspAsnAlaPheValValProGlnGlnAlaValThyArgGlyAla 324
Db 247 AAAAAGCAAACTTAATCAACGATTAATTAATTCGCAAAACGATTAACCGCGG 188
Qy 325 LysAsp-----ThyValMetLeuValAsnAlaGlnGlyValMetGlnProArgGlyVal 342
Db 187 CGTAACGATCAACCGCTCAATTAATTAATTAACGCAATTAACCAATTAACCAACCGCTGATC 128
Qy 343 ThyValAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 362
Db 127 GTTCAACCAACCTTAATTAACGATTAATTAATTAATTAATTAATTAATTAATTAATTA 68
Qy 363 LysValValValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 382

Db 67 CGCGTAATAATAATACTA-----CAAAATAATCGTCT 32
 RESULT 34
 ID ABQ24965 standard; DNA; 607 BP.
 AC ABQ24965;
 DT 12-JUN-2002 (first entry)
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 11556.
 KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;
 KW SNP; cell differentiation; ds.
 OS Homo sapiens.
 XX MO200218632-A2.
 PN 07-MAR-2002.
 PD 01-SEP-2001; 2001WO-EP10074.
 PF 01-SEP-2000; 2000DE-1043826.
 PR 05-SEP-2000; 2000DE-1044543.
 XX (EPIC-) EPIGENOMICS AG.
 PA Olek A, Piepenbrock C, Berlin K, Guetig D;
 PI WPI; 2002-371829/40.
 DR Determining the degree of cytosine methylation in genomic DNA, useful
 PT for diagnosis and prognosis, comprises selective hybridization of
 PT amplicons from chemically treated DNA
 XX
 PS Claim 12; 56pp + Sequence Listing; 56pp; German.
 XX
 CC This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
 CC and the degree of hybridisation to both classes is determined from the
 CC label on the amplicon. From the ratio of labels hybridised to the two
 CC classes of oligomers, the degree of methylation is calculated. The method
 CC is used: (i) for diagnosis and/or prognosis of side effects of
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory
 CC systems etc., particularly by detecting mutations or single nucleotide
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
 CC types and for investigating cell differentiation. The method allows the
 CC methylation status of many C residues to be determined simultaneously.
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
 CC method for determining the degree of cytosine methylation described in
 CC the disclosure of the invention.
 CC
 XX
 SQ Sequence 607 BP; 274 A; 159 C; 53 G; 121 T; 0 other;
 Alignment Scores:
 Pred. No.: 2.24e-15 Length: 607
 Score: 292.00 Matches: 60
 Percent Similarity: 56.00% Conservative: 52
 Best Local Similarity: 30.00% Mismatches: 78
 Query Match: 14.46% Indels: 10
 DB: 24 Gaps: 2
 US-09-889-756a-2 (1-412) x ABQ24965 (1-607)

Qy 185 AlaProIleSerGlyPheIleGlyGlnSerIleValSerGlyGlyThrLeuLeuAsnIa 204
 Db 1 TCTCCGATTACAGATCCATTAAATGACAGCTAACGAAACGATTAAATACAAAC 60
 Qy 205 GlyAspThrThrValLeuAlaThrIleArgGlnThrAsnProMetCylValAsnValThr 224
 Db 61 GATCAACGACCTACGCTAACACCGTACCAACTTAACTGATCTACGTTAAATATAC 120
 Qy 225 GlnSerIleSerGlyValMetIleValArgArgGlnIleAlaGlyIleValLeuAla 244
 Db 121 CAATCCACACAGACCTCTTCCCTTAACGCTTAACAAACAACTAACGATACACGCTAAACAA 180
 Qy 245 AlaAspGlyValIleAlaValGlyIleValPheAspAspGlyThrValTyProGlyIle 264
 Db 181 AAAAAGACAAACAAATAATATCATATACCAATATACCAATATACCAATATATATATATAT 240
 Qy 265 GlyArgLeuLeuPheAlaAspProValValAsnGlySerThrGlyGlnIleThrLeuArg 284
 Db 241 GATACGCTAAATATCTCTTAACGTTACCGCTTAATCAACCACTAAATCTATCACCCTAC 300
 Qy 285 AlaAlaValProAsnAspGlnAsnIleLeuMetProGlyLeuTyThrValArgValLeuMet 304
 Db 301 GCTATCTTCCGAAACCGGAATCACTCTACTACCGAATATATTCGATACGCGACGCTCTA 360
 Qy 305 AspGlnValAlaValAspAsnAlaPheValValProGlnGlnAlaValThrArgIleVal 324
 Db 361 AAAAAAACTTATATCCAAACGCTATTTATATCCGCAACAAACGTAACCGGACCGC 420
 Qy 325 LysAsp-----ThrValMetIleValAsnAlaGlnGlyIleMetGluProArgIleVal 342
 Db 421 CGTACGATACCAACCGTACTTAATTAATTAACGCAATTAACAAATTAACAAACCGTCCGATC 480
 Qy 343 ThrValAlaGlnGlnGlnGlnIleThrAsnTrpIleValThrSerGlyLeuValAspGlyAsp 362
 Db 481 GTTACAAACCAACTATTAACGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 540
 Qy 363 LysValValValGlyGlyIleSerIleAlaGlyIleThrGlyAlaLysValThrPro 382
 Db 541 CGCGTAATAATAATACTA-----CAAAATAATCGTCT 576
 RESULT 35
 AAT42063
 ID AAT42063 standard; DNA; 1830121 BP.
 XX
 AC AAT42063;
 XX
 DT 14-SEP-1999 (first entry)
 XX
 DE Haemophilus influenzae complete genome sequence.
 KW Genome; bacterium; Haemophilus influenzae; computer readable medium;
 KW expression modulating fragment; regulation; gene expression; vector;
 KW organism; open reading frame; ORF; ds.
 OS Haemophilus influenzae.
 XX
 PN WO9633276-A1.
 PD 24-OCT-1996.
 XX
 PF 22-APR-1996; 96MO-US05320.
 XX
 PR 07-JUN-1995; 95US-0487429.
 PR 21-APR-1995; 95US-0426787.
 PR 07-JUN-1995; 95US-0476102.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (UYJO) UNIV JOHNS HOPKINS.
 PI Adams MD, Fleischmann RD, Smith HO, Venter JC, White O;
 DR WPI; 1996-485782/48.

XX Haemophilus influenzae Rd genome recorded on computer readable
 PT medium - useful for identifying commercially important nucleic acid
 PT fragments by homology searching

Claim 1, Page 77.2-77.1091, 1291pp; English.

CC This sequence represents the complete genome sequence of the bacterium
 CC Haemophilus influenzae strain Rd. The invention relates to a computer
 CC readable medium (CRM) having recorded upon it the complete H. influenzae
 CC nucleotide sequence (1), a representative fragment of (1) or a nucleotide
 CC sequence at least 99% identical to (1). By providing the full-length
 CC genomic sequence in a computer readable form, it is possible to identify
 CC commercially important nucleic acid fragments and expression modulating
 CC fragments (EMFs) of the Haemophilus genome. The EMFs can be used to
 CC regulate the expression of a nucleic acid molecule. Vectors and altered
 CC organisms comprising the predicted ORFs can be used to produce any of the
 CC polypeptide fragments of the H. influenzae Rd genome.

XX Sequence 1830121 BP; 567399 A; 350615 C; 347389 G; 564036 T; 682 other;

Alignment Scores:

Pred. No.:	1,14e-10	Length:	1830121
Score:	286.00	Matches:	105
Percent Similarity:	41.01%	Conservative:	66
Best Local Similarity:	25.18%	Mismatches:	148
Query Match:	14.17%	Indels:	98
DB:	17	Gaps:	14

US-09-889-756a-2 (1-412) x AAT42063 (1-1830121)

Qy 6 PhelysalawetargalalalaaleuallalalavalalaleuallleuserSerCys 25

Db 946822 TTTAATATGATTAAGGCGTAATGATTAAGCGGCAATTTGA----- 946863

Qy 26 Glylyseglyaspallalaglnlyglyclnproalaglyalaglyalalproalpro 45

Db 946864 -----GGAATGCCA-----GATCTTCAGGCCCA 946887

Qy 46 ValValGlyValValThrValHisProGlnThrValAlaLeuThrValGlnLeuProGly 65

Db 946888 GTAACC---GCACTTGAAGTCAACGCGTAATGATTAAGCGGCAATTTGA----- 946944

Qy 66 ArgleuGlnSerleuArgThrAlaAspValArgAlaGlnValGlyGlyIleIleGlnLys 85

Db 946945 CTTGTGGTCCCAATCAAGGCGCATGCTCAGTACACAAATGCGCGCGTTTCACAA 947004

Qy 86 ArgleuPheGlnGlnGlySerTyrrValArgAlaGlnProleuTyrrGlnIleAspSer 105

Db 947005 GTACTTGTCAAAATGCAAAATGTGAAAAAGTGAGGCTTGTGAGCTTGATAGT 947064

Qy 106 SerThrTyrrGlnAlaGlnLeuGlnSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThr 125

Db 947065 TCTGTGCAACGAGTAACTTCAACGCTGCTCAGGCAAAATTTATGACACTTCGCAAACT 947124

Qy 126 LeuAlaIalysAlaAspAlaAspLeuAlaArgTyrrProleuValAlaIalaglyAlaVal 145

Db 947125 -----TACCAACGTTATGCTGTTTATTAATAGCAATGCTGTA 947163

Qy 146 SerArgGlnGlnTyrrAspAlaIalValThrAlaIalysArgSerAlaGlnAlaGlyValLys 165

Db 947164 TCACGTCMAAGAAATGATTAACGCAAAAGCGCTTATGATCTCAAGTACGATGATGAA 947223

Qy 166 AlaIalaglnAlaIalIalIalysSerAlaGlyIleAsnLeuAsnArgSerArgIleThrAla 185

Db 947224 TCTCTAATAACGCAATTTGA-----CGTCGTAATAATGTTGCG 947262

Qy 186 ProIleSerGlyPheIleGlyGlnSerTyrrValSerGlnGlyThrLeuLeuAsnAlaGly 205

Db 947263 CCAATTGATGCAAGACAGATATTTGAAAATCAATTTGCAATTTGCAATTTGGA 947322

Qy 206 AspThrThrValleuAlaThrIleArgGlnThrAsnProMetTyrrValAlaValThnGln 225

Db 947323 -----ACAGAAATTTGGCGTGAAGAAATACATAGCTCAATGAAAGCGAATTTGCTCTT 947376

Qy 226 SerAlaSerGlnValMetLysLeuArgGlnIleAlaGlnGlyLysLeuLeuAlaIa 245

Db 947377 TCACAAATATGATTTAGTAAATTTATCAT-----ATCGGTCAAGCGGTTTACAGCG 947424

Qy 246 AspGlyValIleAlaValGlyIleLysPheAspAspGlyThrValTyrrProGlnLysGly 265

Db 947425 ACAACAGATCTCGCTTGGCGGAAACATTT-----TCAGCT 947460

Qy 266 ArgleuLeuPheAlaAspProValAlaAsnGlnSerThrGlyGlnIleThrLeuArgAla 285

Db 947461 CGAATCACTCGCATTTGAACTGCCATTAATTCATCAACAGGTTTATGATGTTGACAGCT 947520

Qy 286 AlaVal---ProAsnAspGlnAsnIleLeuMetProGlyLeuTyrrValArgValLeuMet 304

Db 947521 ACTTTTATCTTGAAGATGGCAATTAATTTCTTCAGGATATGTTCTTCCGTTTACGATT 947580

Qy 305 AspGlnValAlaValAspAsnAlaPheValAlaProGlnGlnAlaValThr----- 321

Db 947581 GCACTTCCAACTGAACAAATCAAGTTGCTTCCACAAAGTATTAAGTATTAAGTATG 947640

Qy 321 ----- 321

Db 947641 TATGCGCAAAATTTGCTTATTTACTTGAACCATTAATTCGAAGCAAGAAAAATATGCA 947700

Qy 322 -----ArgGlyAlaLysAspThrValMetIleValAsn 332

Db 947701 GGTAATGAAAATTTGATCGTCTCTATCGTCCGAACAGATCACCGTATTTACTTAAGAT 947760

Qy 333 AlaGlnGlyGlyMetGlnProArgGlnValThrValAlaGlnGlnGlnGlyThrAsnTrp 352

Db 947761 CGTCAAGT-----GTTATGCTCAATTAACAGGAAATGAA--- 947796

Qy 353 IleValThrSerGlyLeuLysAspGlyAspLysValValAlaGlnGlyIleSerIleAla 372

Db 947797 -----GTTAAAGTGGAGATTAATTAATTAACAGCGGCTCAGCAA----- 947835

Qy 373 GlyIleThrGlyAlaLysLysValThrProLysGlnThrAlaSerSerGlu 389

Db 947836 GGTATTGTAATGAAGCTTTGTG-----GAATGATTAATAAAAGAC 947877

RESULT 36

ID ABO90313 standard; DNA; 1086 BP.

AC ABO90313;

DT 01-OCT-2002 (first entry)

DE M. capsulatus gene #298 for DNA array.

XX Micro array; gene; ds; differential expression; gene expression.

OS Methylococcus capsulatus.

PN W0200255655-A2.

PD 18-JUL-2002.

PF 14-JAN-2002; 2002WC-NO00019.

PR 12-JAN-2001; 2001NO-0000235.

PR 12-JAN-2001; 2001NO-0000239.

PA (UNIF-) UNIFOB STIFTELSEN UNIV BERGEN.

XX (TIGR-) TIGR.

PI Birkeland NK, Eidhammer I, Jonassen I, Jensen HB, Lien T;

PI Lillhaug JR, Lossius I, Eisen JA, Frazer CM, Durkin AS;

PI Salzberg SL;

DR WPI; 2002-557818/59.

CC (ORFs) of *Neisseria meningitidis* and *N. gonorrhoeae* which encode
 CC antigenic proteins (see AAY3499-Y38944). The antigenic proteins, their
 CC fragments, their nucleic acids and antibodies are useful for diagnosis,
 CC prevention (as vaccines) or treatment of *Neisseria* infections,
 CC such as meningitis, septicemia and gonorrhea. Both organisms
 CC are closely related. Fragments of the nucleic acids are useful
 CC as hybridisation probes and antisense reagents.

XX Sequence 1005 BP; 271 A; 233 C; 315 G; 186 T; 0 other;

Alignment Scores:

Pred. No.:	2.23e-13	Length:	1005
Score:	271.00	Matches:	100
Percent Similarity:	45.45%	Conservative:	60
Best Local Similarity:	28.41%	Mismatches:	141
Query Match:	13.42%	Indels:	52
DB:	20	Gaps:	10

US-09-889-756a-2 (1-412) x AA212304 (1-1005)

```

Qy 75 ValArgAlaGlnValGlyGlyLeuGlnLeuArgLeuPheGlnGlnGlySerThrVal 94
Db 7 GTGGCGCCGCGAGCGATGCGGCGAGATTAAGTACTTATGTCACAACTCGGCGAAGGTT 66
Qy 95 ArgAlaGlyGlnProLeuThrGlnLeuSerSer-----SerThr 107
Db 67 AAAAAGCGCATTTGTCGGAATCAATTCAGCTCGAGCAACATACGCTCAATACG 126
Qy 108 TyrGlnAlaSerLeuGlnSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAla 127
Db 127 GAAATATCCAGTTGGAAAGCGTATCAGCGAAGCTGTGTGCGACGATTCATTGGGC 186
Qy 128 LysAlaAspAlaAspLeuAlaArgGlyLeuProLeuAlaAlaGlnAlaValSerArg 147
Db 187 AGCGCGAGAGAAATTAACCGTCAGCGCGCTTATGAGAAAGAAACGCACTCCAA 246
Qy 148 GlnGlnTyrAspAlaAlaValThrAlaLysArgSerAlaGlnAlaGlyValLysAlaAla 167
Db 247 GAGGATTTGGAAAGCGCGAGATGCTTTGCCCGCAAGCCATGTTGCCGAGCTG 306
Qy 168 GlnAlaAlaLysSerAlaGlyLeuAsnLeuAsnArg-----180
Db 307 AAGGCTTAATCAAGACAGCAAAATTTCCATCATATCCCGAGTCGAAATGGGCTAC 366
Qy 181 SerArgLeuThrAlaProLeuSerGlyPheLeuGlnSerLysValSerGlnGlyThr 200
Db 367 ACGGCGATTTACCGGAGATGAGACGACGCGTGGCGATTCCTCGTGAAGAGGGGCGAG 426
Qy 201 LeuLeuAsnAlaGlyAspThrThrValLeuAlaThrLeuArgGlnThrAsnProMetThr 220
Db 427 ACTGTGAACCGGCG-----CAGTCTACCGGAGCAT 459
Qy 221 ValAsnValThrGlnSerAlaSerGlnValMetLysLeuArgArgGlnLeuAlaGlnGly 240
Db 460 GTCCAAATTTGCGCAT-----CTGGATATGATGTTGAACAAATGACAGATTCGCGAGGC 513
Qy 241 LysLeu-----LeuAlaAlaAspGlyValLeuAlaValGlyLeuLysPheAspArgGly 258
Db 514 GATATTACCAAGGTGAAGCGGCGGAGAGATTTCTGTTTCGATTTTGTCCGAACCGGAT 573
Qy 259 ThrValTyrProGlnLysGlyArgLeuLeuPheAlaAspProValVal-----274
Db 574 ACG-----CGATTAAAGGAGCTCGACAGCGGTGACCCCGGCGTACACAGATTCG 627
Qy 275 -----AsnGlnSerThrGlyGlnLeuThr-----LeuArg 284
Db 628 TCGGCGCGATTACACAGCGATACGATACGCGCTTCACATCGGCTACTATTATGCCCCGT 687
Qy 285 AlaAlaValProAsnAspGlnAsnLeuLeuMetProGlyLeuThrValArgValLeuMet 304
Db 688 TCGTTGTGCGGAATCCGAGCGGCAACTCGCCACGCGGGGATGACGACGAATACGGTT 747
Qy 305 AspGlnValAlaValAspAsnAlaPheValValProGlnGlnAlaVal---ThrArgGly 323

```

```

Db 748 GAATCGACGCGCGTGAATAATGTGCTGATTATTCCTGCTGACCGTGAAATATCGCGGC 807
Qy 324 AlaLysAspThrValMetLeuValAsnAlaGlnGlyLysGlnPheGlnPheGlnValThr 343
Db 808 GGCAGAGCGCTTTGTGCGCGTGTGGTCCGACGCGCAAGCGCGGCGGCAACCGCAATCCG 867
Qy 344 ValAlaGlnGlnGlnGlnThrAsnTrpLeuValThrSerGlyLeuLysAspGlyAspLys 363
Db 868 ACCGATATGAGACAGATGATGATACGAAATGTAATAAGCGGCTTGAAGAAGGCGGACAAA 927
Qy 364 ValValAlaGlnGlyLysSerLeuAlaGlyLeuThrGlyAlaLysLysValThrProLys 383
Db 928 GTGCTATCTCCGAAATTAACCGCGCGA-GCAACAGAAAGCGGGAACCGCCCTACG 986
Qy 384 GlnTrpAlaSerSerGlnGlnGlnAlaAlaPro 395
Db 987 -----CGGCCCGCGCGCG 1001

RESULT 38
ID AA253632
AA253632 standard; DNA; 1005 BP.
AC AA253632;
AC 21-MAR-2000 (first entry)
DE Neisseria meningitidis ORF 290 partial DNA sequence SEQ ID NO:1213.
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy; ds.
OS Neisseria meningitidis.
XX
XX WO957280-A2.
XX 11-NOV-1999.
XX
XX 30-APR-1999; 99WO-US09346.
XX
XX 01-MAY-1998; 98US-0083758.
XX 31-JUL-1998; 98US-0094869.
XX 02-SEP-1998; 98US-0098994.
XX 02-SEP-1998; 98US-0099062.
XX 09-OCT-1998; 98US-0103749.
XX 09-OCT-1998; 98US-0103794.
XX 09-OCT-1998; 98US-0103796.
XX 25-FEB-1999; 99US-0121528.
XX
XX (CHIR) CHIRON CORP.
XX (GENO-) INST GENOMIC RES.
XX
XX Fraser C, Galicotti C, Grandi G, Hickey E, Masignani V, Mora M,
XX Petersen J, Pizza M, Rapunoli R, Ratti G, Scalato E, Scarselli M,
XX Tettelin H, Venter JC;
XX
XX WPI; 2000-062150/05.
XX P-PSDB; AAY74870.
XX
XX Novel Neisseria polypeptides predicted to be useful antigens for
XX vaccines and diagnostics
XX
XX Claim 7; Page 676-677; 1453pp; English.
XX
XX AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941
XX represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
XX and polypeptides. AA254576 and AA254616 to AA255473 represent
XX PCR primers used in the exemplification of the present invention. The
XX polypeptides, the polynucleotides, antibodies and compositions of
XX the invention can be used as vaccines, as diagnostic reagents, and as
XX immunogenic compositions. The polypeptides can be used in the
XX manufacture of medicaments for treating or preventing infection due to

```

CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.

XX Sequence 1005 BP, 271 A, 233 C, 315 G, 186 T, 0 other;

Alignment Scores:

Pred. No.:	2,23e-13	Length:	1005
Score:	271.00	Matches:	100
Percent Similarity:	45.45%	Conservative:	60
Best Local Similarity:	28.41%	Mismatches:	141
Query Match:	13.42%	Indels:	52
		Gaps:	10

US-09-889-756a-2 (1-412) x AA53632 (1-1005)

```

Qy 75 ValAlaGAlaGlnValGlyGlyLeuGlnLysArgLeuPheGlnGlySerTyrVal 94
Db 7 GTCGGCGCGCAGCATCGGCGCAGATTAAGATTAATCTTATCTCAACTCGGCAACAGGTT 66
Qy 95 ArgAlaGlyGlnProLeuTyrGlnLeuSer-----SerThr 107
Db 67 AAAAAGGCGATTGTCGGAATCAATTCGACCTCGACCAATACCTCGAATACG 126
Qy 108 TyrGlnAlaAsnLeuGlnSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAla 127
Db 127 GAAAAATCCAGTTCGAAACGATACGAGCGAAGCTGCTGCGCAGATTCGATTCGGC 186
Qy 128 LysAlaAspAlaAspLeuAlaArgTyrLysProLeuValAlaAlaGlnAlaValSerArg 147
Db 187 AGCCGCGGAGAAATAATATACGTCAGCGCGGCTATGAGAAAGAAAGCGCATTCGAAA 246
Qy 148 GlnGlnTyrAspAlaAlaValAlaThrAlaLysArgSerAlaGlnAlaGlyValLysAlaAla 167
Db 247 GAGGATTGGAAGCGCGCAGATGCTGTTGCCCGCGCAAGCCAAATGTTGCCGAGCTG 306
Qy 168 GlnAlaAlaLeuLysSerAlaGlyLeuAsnLeuAsnArg----- 180
Db 307 AAGGCTTAACTCAGACGACGAAATTTCCATCAATACCGCGCAGTGGGAATTGGGCTAC 366
Qy 181 SerArgLThrAlaProLysSerGlyPheLleGlnSerLysValSerGlnGlyThr 200
Db 367 ACGGCGATTACCGCAACGATGACGACGCGTGGCGGATTCGTCGAGAGAGGCGAG 426
Qy 201 LeuLeuAsnAlaGlyAspThrThrValLeuAlaThrLysArgLThrAspProMetTyr 220
Db 427 ACTGTGAACGCGCGCG-----CAGTCAACCGCCGACGATT 459
Qy 221 ValAsnValThrGlnSerAlaSerGlnValMetLysLeuAlaArgGlnAlaGlnGly 240
Db 460 GTCCAAATGGCGAAT-----CTGGATATGATGTTGAACAAATATGCGATTCGAGGCG 513
Qy 241 LysLeu-----LeuAlaAlaAspGlyValLleAlaValGlyLysPheAspGly 258
Db 514 GATATTCACCAAGGAGGAGCGCGGCGAGATATTCGTTACGATTTGTCGAAACCGGAT 573
Qy 259 ThrValTyrProGlnLysGlyArgLeuLeuPheAlaAspProValAla----- 274
Db 574 ACG-----CCGATTAAAGCGCAACCTCGACGCGTCCGCGGCTGACCAAGATGCG 627
Qy 275 -----AangLysThrGlyGlnLThr-----LeuArg 284
Db 628 TCGGCGCGTTACACAGCATGACGATACGCGCTTCAATGCGGCTACTATTATGCGCGT 687
Qy 285 AlaAlaValProAsnAspGlnAsnLleLeuMetProGlyLeuTyrValAlaArgValLeuMet 304
Db 688 TCGTTTGGCCGAATTCGCGAAGCAACTCGCCACGCGGAGATGACGACGACAAATACGTT 747
Qy 305 AspGlnValAlaValAspAsnAlaPheValValProGlnGlnAlaVal---ThrArgGly 323
Db 748 GAAATCGACGCGCGTGAATAATGTCTGATATTCGTCGTCGACCGTGAAATAATCGCGCG 807
  
```

```

Qy 324 AlaLysAspThrValMetLleValAsnAlaGlnGlyMetGluProArgGluValThr 343
Db 808 GCGAAGCGCTTTGGCCGCTTGGTCGTCGCGACGCGCAAGCGCGCAACCGAATATCCG 867
Qy 344 ValAlaGlnGlnGlnGlyThrAsnTyrLleValThrSerGlyLeuLysAspGlyAspLys 363
Db 868 ACCGCTATGAGACAGCAGTATGAATACCGAAGTAAAGCCGCTTGAAGAGCGGACAAA 927
Qy 364 ValValValAlaGlnGlyLysSerLleAlaGlyLleThrGlyAlaLysLysValThrProLys 383
Db 928 GTGTCATCTCCGAAATACCGCGCCGCA-GCAACAGAGAAAGCGCGCAAGCGCCCTTAG 986
Qy 384 GluTyrAlaSerSerGluAsnGlnAlaAlaAlaPro 395
Db 987 -----CGGCGCGCGCGCGCG 1001

RESULT 39
ABQ91605
ID ABQ91605 standard; DNA, 900 BP.
XX
AC ABQ91605:
XX
DT 01-OCT-2002 (first entry)
XX
DE M. capsulatus gene #1590 for DNA array.
XX
KW Micro array; gene; ds; differential expression; gene expression.
XX
OS Methylococcus capsulatus.
XX
PN W020255655-A2.
XX
PD 18-JUL-2002.
XX
PF 14-JAN-2002; 2002W0-NO00019.
XX
PR 12-JAN-2001; 2001NO-0000235.
XX
PT 12-JAN-2001; 2001NO-0000239.
XX
PA (UNIF-) UNIFOB STIFTELSEN UNIV BERGEN.
XX
PE (TIGR-) TIGR.
XX
PI BitKeland NK, Eidhammer I, Jonassen I, Jensen HB, Lien T,
PI Lillehaug JR, Lossius I, Eisen JA, Fraser CM, Durkin AS,
PI Salzberg SL.
XX
XX WPI; 2002-557818/59.
XX
XX
XX Novel DNA array useful for determining differential expression of
XX Methylococcus capsulatus genes, comprises polynucleotides or
XX oligonucleotides representative for a selective number of Methylococcus
XX capsulatus genes -
XX
XX Claim 14; Page 607; 678bp; English.
XX
XX The invention relates to a novel DNA array giving a representation of a
XX number of Methylococcus capsulatus genes. The method of the invention is
XX useful for determination of the differential expression of the genes of
XX M. capsulatus, and for studying gene expression on a genomic scale and in
XX gene expression assays of M. capsulatus genes. The sequences shown in
XX ABQ90016-ABQ91855 represent M. capsulatus genes for use in arrays of the
XX invention.
XX
XX
SQ Sequence 900 BP, 137 A, 308 C, 345 G, 110 T, 0 other;

Alignment Scores:
Pred. No.: 7.79e-11 Length: 900
Score: 239.50 Matches: 97
Percent Similarity: 47.71% Conservative: 49
Best Local Similarity: 31.70% Mismatches: 128
Query Match: 11.86% Indels: 33
DB: 24 Gaps: 10
  
```



```
Oy 227 AlaSerGluValMetLysLeuArgGlnIleAlaGlnGlyLysLeuAlaIleAsp 246
      ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
Db 127 AGTAACGATTTTTCGCTTGAATAGCAATTGGCAATGCTTGAATAAAGAGAAC 186
      ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
Oy 247 GlyValIleAlaValGlyIleLysPheAspGlyThrValTyrProGlnLysGlyArg 266
      ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
Db 187 GGTAAAGTTAAAGTGTATGTATTAAGTAGCGGTATTAAGTTTCGTAGGACCGTACG 246
      ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
Oy 267 LeuLeuPheAlaAspProValValAsnGlnUserThrGlyGlnIleThrLeuArgAlaIle 286
      ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
Db 247 TTGCAATTTTTCGACGTTATCGCTTATAGATTATGGGTTTATTATTACGCGTTATT 306
      ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
Oy 287 ValProAsnAspGlnAsnIleLeuMetProGlyLeuTyrValArgValIleuMetAspGln 306
      ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
Db 307 TTTTCGATTCGCAATTATTTTGTGTGCGGTATGTCGTGCGGTACGTTTGAAGAA 366
      ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
Oy 307 ValAlaValAspAsnAlaPheValValProGlnGlnAlaValThrArgGlyAlaLysAsp 326
      ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
Db 367 GGGTTTAATTAAAGTTATTTTAACTTCGTAATAGGCGGTAAATTCGTACGTCGCGTGC 426
      ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
Oy 327 -----ThrValMetIleValAsnAlaGlnGlyMetGluProArgGluValThrVal 344
      ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
Db 427 GATGTTATCGTATTCGTATGTCGCGCGCATGATAAGTGAATTCGTCGATCGTTGTA 486
      ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
Oy 345 AlaGlnGlnGlyThrAsnTrpIleValThrSerGlyLeuLysAspGlyAspLysVal 364
      ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
Db 487 AGTTAGGTTATTCGCGCATAGGTCGTCGATAGAAAGTTGAAGTAGCGCATCGCCTA 546
      ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
Oy 365 ValValGlnGlyIle 369
      ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
Db 547 GTAATTAAGTGGGTTG 561
      ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
```

Search completed: September 8, 2003, 06:03:16
Job time : 3427 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 8, 2003, 04:58:37 ; Search time 4846 Seconds
(without alignments)
3478.078 Million cell updates/sec

Title: US-09-889-756A-2
Sequence: 1 MAFYAFKRAMRAALAAVAL.....AAPOSGVQTASEAKTASAE 412

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgm2.1/USPTO.spool/US09889756/runat.05092003.124324.23279/app.query.fasta.1.583
-DB=GenEmbl -QMT=fastcap -SUFFIX=rge -MINMATCH=0.1 -LOOPCU=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40 -MODE=LOCAL
-OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09889756 @CGN 1.1 3508 @runat.05092003.124324.23279 -NCPU=6 -ICPU=3
-NO MAP -LARGESUBRY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: GenEmbl:*
2: gb_hhg:*
3: gb_in:*
4: gb_on:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pt:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vt:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vt:*
30: em_hhg_hum:*
31: em_hhg_inv:*
32: em_hhg_other:*
33: em_hhg_mus:*
34: em_hhg_pln:*
35: em_hhg_rod:*
36: em_hhg_mam:*
37: em_hhg_vrt:*
38: em_sy:*
39: em_hhgo_hum:*
40: em_hhgo_mus:*
41: em_hhgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	# Match	Length	DB ID	Description
1	2019	100.0	1239	6 AX027848	AX027848 Sequence
2	1993	98.7	326301	1 NMA622491	AL62757 Neisseria
3	1989	98.5	1239	6 AX391542	AX391542 Sequence
4	1989	98.5	14056	1 AE002521	AE002521 Neisseria
5	1989	98.5	349980	6 AX044033	AX044033 Sequence
6	1938	96.0	1500	1 NGU14993	U14993 Neisseria g
7	1908.5	94.5	1235	1 AF037041	AF037041 Neisseria
8	1278	63.3	2127	1 NGMT6RC	N25796 N.gonorrhoea
9	839	41.6	215050	1 AL646057	AL646057 Ralstonia
10	777	38.5	20382	1 AE008856	AE008856 Salmone
11	770	38.1	11110	1 AE005757	AE005757 Caulobact
12	765	37.9	13945	1 AE012380	AE012380 Xanthomon
13	763.5	37.8	6066	1 EAE306389	AJ306389 Enterobac
14	762.5	37.8	4879	6 E27981	M94248 E. coli acr
15	762.5	37.8	4879	6 E27981	E27981 Organic soi
16	762.5	37.8	6039	1 ECU00734	U00734 Escherichia
17	762.5	37.8	12524	1 AE005225	AE005225 Escherich
18	762.5	37.8	13446	1 AE000152	AE000152 Escherich
19	762.5	37.8	159818	1 ECU82664	U82664 Escherichia
20	762.5	37.8	295741	1 AP002551	AP002551 Escherich
21	760.5	37.7	300817	1 AE016756	AE016756 Escherich
22	759.5	37.6	11927	1 AE015072	AE015072 Shigella
23	759.5	37.6	292309	1 AE016979	AE016979 Shigella
24	758.5	37.6	5281	1 AY061647	AY061647 Proteus m
25	758	37.5	197050	1 AL646081	AL646081 Ralstonia
26	756	37.4	240050	1 AL627267	AL627267 Salmone
27	756	37.4	300029	1 AE016842	AE016842 Salmone
28	753	37.3	5600	1 PSEENVCD	L11616 Pseudomonas
29	753	37.3	13987	1 AE004479	AE004479 Pseudomon
30	753	37.3	20347	1 AE008717	AE008717 Salmone
31	750.5	37.2	4571	1 ECOACREB	M96848 E. coli acr
32	750.5	37.2	6803	1 ECENVCD	X57948 E.coli envc
33	750.5	37.2	11095	1 AE000405	AE000405 Escherich
34	750.5	37.2	110000	1 ECOLUM67.1	Continuation (2 of
35	749.5	37.1	11807	1 AE005554	AE005554 Escherich
36	749.5	37.1	307962	1 AP002564	AP002564 Escherich
37	747	37.0	5914	1 SMA252200	AP002564 Escherich
38	745.5	36.9	300359	1 AE016767	AJ252200 Stenotrop
39	741.5	36.7	12478	1 AE011925	AE011925 Escherich
40	738	36.6	4503	1 AB104882	AB104882 Xanthomon
41	735.5	36.4	301995	1 AE016779	AB104882 Xanthomon
42	734	36.4	11048	1 AE015901	AE015901 Pseudomon
43	732	36.3	110000	2 AC141230_0	AE015901 Shewanell
44	731	36.2	254677	2 AC068494	AC141230 Homo sapi
45	730.5	36.2	12296	1 AF031417	AC068494 Mus muscu
					AF031417 Pseudomon

RESULT 1

ALIGNMENTS

AX027848
 LOCUS AX027848 1239 bp DNA linear PAT 24-NOV-2000
 DEFINITION Sequence 1 from Parent W00043517.
 ACCESSION AX027848 AX027853
 VERSION AX027848.1 GI:10188692
 KEYWORDS
 ORGANISM *Neisseria meningitidis*
 SOURCE *Neisseria meningitidis*
 Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 Neisseriaceae; Neisseria.
 REFERENCE
 1. Thonard, J.
 AUTHORs Baab055 polynucleotide and polypeptide from *neisseria meningitidis*.
 TITLE Uses thereof
 JOURNAL Patent: WO 0043517-A 1 27-JUL-2000;
 COMMENT SMITHKLINE BEECHAM BIOLOG (BE) ; THONARD JOELLE (BE)
 FEATURES On Oct 15, 2002 this sequence version replaced gi:10188695.
 source
 1. 1239
 /organism="Neisseria meningitidis"
 /mol_type="genomic DNA"
 /db_xref="taxon:487"
 BASE COUNT 275 a 332 c 384 g 248 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 3,21e-112 Length: 1239
 Score: 2019.00 Matches: 412
 Percent Similarity: 100.00% Conserved: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 Gaps: 0
 DB: 6
 US-09-889-756a-2 (1-412) x AX027848 (1-1239)

541 TCGGCATTACCGCCGCGATTTCCGGCTTATCGGTACGTCACAAAGTTTCCGAAGTACG 600
 Db
 201 LeuLeuSnaIaGlaIaAaPThrThrValIleuAlaThrIleArgIlnTrAspProMetTyr 220
 Oy
 601 CTGTTGAATGGGGCGATACACCGCTCTGGCAACCATCCGCCAACAACCATCCGATGAT 660
 Db
 221 ValAsnValThrgInSerIaSerGluValMetLysLeuArgArgIlnIleAGluGly 240
 Oy
 661 GTGAACGTTACCCAGCTCGATCCGAAGTATGAAATTGGCGCGCTCAGATAGCGAAGGC 720
 Db
 241 LysLeuLeuAlaAaPglyValIleAlaValGlyIleLysPheAspAspGlyThrVal 260
 Oy
 721 AAACGTGTGGGGGGGATGGTGTATTCGGTCCGATCAAAATTGACAGAGCGACAGTT 780
 Db
 261 TyrProGluLysGlyAArgLeuPheAlaAspProValValaGluSerThrGlyGln 280
 Oy
 781 TACCTTAAAAAGGCCCTCTCTCTCTTGGCCGATCCGCTCGTCAACGATGACCGGTACG 840
 Db
 281 IleThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetProGlyLeuTyrVal 300
 Oy
 841 ATTACCTGGCGCGCGCGCGATACCGAACGATCAGAAATCCGTGATGCGCGCTGTATGTG 900
 Db
 301 ArgValIleuMetAspGlnValAlaValAspAsnAlaPheValValProGlnGlnAlaVal 320
 Oy
 901 CGCGTGTGATGACCAAGTGGCGGTGATTAACCATTTGTGTGCGCAGACGAGCGGTA 960
 Db
 321 ThrArgGlyAlaLysAspThrValMetIleValAsnAlaGlnGlyGlyMetGluProArg 340
 Oy
 961 ACCGCGGTGGCAAGATACCGTATATGTGATATCCCAAGCGGTATGGAACCCCGC 1020
 Db
 341 GluValThrValAlaGlnGlnGlnGlyThrAsnTrpIleValThrSerGlyLeuLysAsp 360
 Oy
 1021 GAGGTAAAGGTTGGCAACAGACAGGTACGATGATGATTTGATTCGTGGGTCTGAAGAC 1080
 Db
 361 GlyAspLysValValValGluGlyIleSerIleIleAlaGlyIleThrGlyAlaLysVal 380
 Oy
 1081 GGGGACAAAGGTGTGTGGAAGCATCAGATCCCGGTATACGGGTGCACAAAAGGTA 1140
 Db
 381 ThrProLysGluTrpAlaSerSerGluAsnGlnAlaAlaAlaProGlnSerGlyValGln 400
 Oy
 1141 ACCGCCAAGATGGGGTGTCTCTGAACCAACGACCGCGGCTCTCAATCCGGGCTTCAG 1200
 Db
 401 ThrAlaSerGluAlaLysThrAlaSerGluAlaGlu 412
 Oy
 1201 ACCGCATCTGAAGCCAAACGCTTTCGAAGCGAA 1236
 Db
 RESULT 2
 NMA622491/c 326301 bp DNA linear BCT 02-SEP-2002
 LOCUS NMA622491
 DEFINITION *Neisseria meningitidis* serogroup A strain Z2491 complete genome;
 segment 6/7.
 ACCESSION AL162757 AL157959
 VERSION AL162757.2 GI:7380371
 KEYWORDS
 ORGANISM *Neisseria meningitidis* Z2491
 SOURCE *Neisseria meningitidis* Z2491
 Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 Neisseriaceae; Neisseria.
 REFERENCE
 1 (bases 1 to 326301)
 AUTHORs Parkhill, J., Achtman, M., James, K. D., Bentley, S. D., Churcher, C.,
 Klee, S. R., Morelli, G., Basham, D., Brown, D., Chillingworth, T.,
 Davies, R. M., Davis, P., Devlin, K., Feltwell, T., Hamlin, N.,
 Holtroyd, S., Jagels, K., Leather, S., Moule, S., Mungall, K.,
 Quail, M. A., Rajandream, M. A., Rutherford, K. M., Simmonds, M.,
 Skellern, J., Whitehead, S., Spratt, B. G. and Barrall, B. G.
 TITLE Complete DNA sequence of a serogroup A strain of *Neisseria*
 meningitidis Z2491
 JOURNAL Nature 404 (6777), 502-506 (2000)
 MEDLINE 20222556
 PUBMED 10761919
 REFERENCE
 2 (bases 1 to 326301)
 AUTHORs Parkhill, J.
 TITLE Direct Submission

/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
/complement(4486..4495)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
gene
4570..7320
/gene="seca"
4570..7320
/gene="seca"
/note="NMA1735, seca, preprotein translocase SecA subunit,
len: 916aa; similar to many eg. SW:P10408 (SECA_ECOLI)
preprotein translocase SecA subunit from Escherichia coli
(901 aa) fasta scores: E: 0, 58.0% identity in 920 aa
overlap. Contains Pfam match to entry PF01043
SecA protein, SecA protein, amino terminal region and
Prosite match to PS01312 Protein seca signatures."
/codon_start=1
/transl_table=1
/product="preprotein translocase SecA subunit"
/protein_id="CAB84963.1"
/db_xref="GI:7380377"
/db_xref="SPTREMBL:Q9UTK7"
/translation="MLTNIAKKIFGSRNDLLKQKRSVARNALNEEQMALSPADLQ
AKTAEFKORLADQGTLDGILPEAPVCEASRRLGMRHPVOLIGKAVLHDGTAEM
RTGEGKLVATLAVLNLAKGIVAVVNDYLSRDMGMEPLVNFGLDGVYIISD
MOPDRONAVAADITVGNNEFGDYLRDNNVTOYDVRSELPKAVVDEVSILIDE
ARPLITISGADNNIOLOVIMNTVPPHLYROETSEEDYVDSKAHOVLISEAGHH
AEQILTMGLAENDSLYSANIALMHMLAALRAHFLPHDQYTVIODGIVYDER
TGRLMSGRWSEGLHQAVEAKEVEIKRENOLTLSTFONYFLYTLKSGMTGADTE

Alignment Scores:

Pred. No.: 1.67e-107 Length: 326301
Score: 1993.00 Matches: 406
Percent Similarity: 99.03% Conservative: 2
Best Local Similarity: 98.54% Mismatches: 4
Query Match: 98.71% Indels: 0
Gaps: 0

US-09-889-756a-2 (1-412) x NMA622491 (1-326301)

QY 1 MetAlaPheTYrAlaPheLYaLaMeCaRGAAlaAlaAlaLeuAlaAlaValAlaLeu 20
Db 245789 ATGGCTTTTAAAGCTTTTAAAGCGCATGCGTGGCGCGCTGGCTGGCGCGCATTTG 245730
QY 21 ValLeuSerSerCYsGLyLysGLyGLyAspAlaAlaGlnGlyGlnProAlaGlyArg 40
Db 245729 GTACTGCTGCTCTGGGTAAGCGGAGACCGGCGCGGCGGCGGCGGCTGCTGCTGG 245670
QY 41 GluAlaProAlaProValAlaGlyValAlaThrValHisProGlnThrValAlaLeuThr 60
Db 245669 GAAGCCCTTGCGCGCTGCTGCTTAACCGTCATCCGCAAAACCGTGCATTTGACC 245610
QY 61 ValGluLeuProGlyArgLeuGlnSerLeuArgThrAlaAspValArgAlaGlnValGly 80
Db 245609 GTCCAGTTGCGGCGGCGCTTGAATCGCTGCGTACCGCGCATGTCGCCCAAGTCGCGC 245550
QY 81 GlyIleIleGlnLysArgLeuPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 100
Db 245549 GGCAATCATCAAAAGCGCTGTTCCAGAGGCGATGATGTCGCGCGGAGACGCGCTG 245490
QY 101 TyrGlnIleAspSerSerThrTyrGlnAlaAsnLeuGlnSerAlaArgAlaGlnLeuAla 120
Db 245489 TATCAGATCGACAGTTTCACTTATGACAGGTCTCGAAAGTGGCGCGCGCAACTGCA 245430
QY 121 ThrAlaGlnAlaThrLeuAlaLysAlaAspAlaAspLeuAlaArgTyrLysProLeuVal 140
Db 245429 ACGGCTCAGCAACCGCTTGCACAAAGCGATGCGGATTTGGCGCGGTACAAACCTTTGGT 245370
QY 141 AlaAlaGlnAlaValSerArgGlnGlnLysArgAlaAlaValThrAlaLysArgSerAla 160
Db 245369 GCCCGCAAGCGCTGACGCGGAGAAATACATGCTGCGGTAACGCGCAACGTTCTGCC 245310
QY 161 GluAlaGlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsnArg 180

Db 245309 GAGGCAAGTGTCAAAAGCAGACAGGCGGCATCAAAATCTCGGCATCAATCTGAACCGT 245250
QY 181 SerArgIleThrAlaProIleSerGlyPheIleGlyIleSerLysValSerGlnGlyThr 200
Db 245249 TCGGCAATTACCGCGCGGATTTTCGGCTTTATCGGTAGTCGCAAAAGTTTCCGAAGTACG 245190
QY 201 LeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAsnProMetTyr 220
Db 245189 CTGTGGAATCGCGCGCATAGACCGCTCTGGCAACCATCCGCCAAACCATCCGATGAT 245130
QY 221 ValAsnValThrGlnSerAlaSerGlnValMetLysLeuArgArgGlnIleAlaGlnGly 240
Db 245129 GTGAACGTTACCCAGCTCTGCATCCGAAGTATGAAATTTGCCGCTCAGATAGCGAAGGC 245070
QY 241 LysLeuLeuAlaAlaAspGlyValIleAlaValGlyIleLysPheAspAspGlyThrVal 260
Db 245069 AACTGCTGGCGGGGAGTGTGTATGCGGTCCGTCATCAAAATTTGACGACGCGCAGTT 245010
QY 261 TyrProGlnLysGlyArgLeuLeuPheAlaAspProValValAsnGlnSerThrGlyGln 280
Db 245009 TACCTTAATAAAGCCGCTCTGCTGTGTTGCCGATCCGACCGTCACGATCGACCGGTACG 244950
QY 281 IleThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetProGlyLeuTyrVal 300
Db 244949 ATTACCTCGCGCGCGCGCGTACCGAACGATCAGAAATATCTGATGCGCGGCTGTATGTG 244890
QY 301 ArgValLeuMetAspGlnValAlaValAspAsnAlaPheValValProGlnGlnAlaVal 320
Db 244889 CGCTGTGTATGACCAAGTGGCGGTGATTAACCATTCATTCGTCGCGCAGCAGCGGTA 244830
QY 321 ThrArgGlyAlaLysAspThrValMetIleValAsnAlaGlnGlyGlyMetGluProArg 340
Db 244829 ACGCGCGGTGCGAAAGTATACCGTATGATGATTCGATCCCAAGCGGTATCGAACCCCGC 244770
QY 341 GluValThrValAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 360
Db 244769 GAGGTAAAGTGTGCGCAACAGCAGGATGATTAACCATTCATTCGTCGCGTGTAAAGGAC 244710
QY 361 GlyAspLysValValAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 380
Db 244709 GGGGACAAAGGTGTGTGGAAGGATCAGTATCCCGGTATTAACGCGGTGCGAAAGGTA 244650
QY 381 ThrProLysGlnTTPAlaSerSerGlnLysGlnAlaAlaAlaAlaProGlnSerGlyValGln 400
Db 244649 ACGGCCAAAGATGGGCGCGCTCGAATAACCAAGCCCGCGCTCAAGCTGACGTTGAG 244590
QY 401 ThrAlaSerGlnAlaLysThrAlaSerGlnAlaGln 412
Db 244589 ACGGCATCTGAAGCCAAACCTGTTGTAAGCGGAA 244554
RESULT 3
AX391542 1239 bp DNA linear PAT 23-MAR-2002
LOCUS AX391542
DEFINITION Sequence 3 from Patent WO0216612.
ACCESSION AX391542
VERSION AX391542.1 GI:19700149
KEYWORDS
SOURCE
ORGANISM
Neisseria meningitidis
Neisseria meningitidis
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
Neisseriaceae; Neisseria.
REFERENCE
1 Lane, J.D., Hughes, M.J. and Santangelo, J.D.
AUTHORS
TITLE Genes and proteins from neisseria meningitidis, and their uses
JOURNAL Patent: WO 0216612-A 3 28-FEB-2002;
Microscience Limited.(GB)
FEATURES
Location/Qualifiers
1..1239
/organism="Neisseria meningitidis"
/mol_type="genomic DNA"
/db_xref="taxon:487"
1..1239
CDS

```

/nc="unnamed protein product"
/codon_start=1
/translation="1"
/protein_id="CAD28957.1"
/db_xref="GI:19700150"
/translation="MAFAFKMRAAALAAVALVLSCKGKGDAAAGCGPAGAEAP
FVGVVTVHPQVLAIVELPGRLESLRTADVAQVGIIGRLFOESYVAGPLYO
IDSTYEGALSARAOALTAQATLAKADADLARXPLVAARVROEYDAVATKRS
EAGYAAQAAIKSAGISLNRRTAPISGFIGOSKSEGTLLNAGDVTVALRTQNP
MYNVTGASAEVMLRROJAEKLLAADVAGIKPDGTVPEKGRLEPADPANE
STGQITLRAVNDONILMPLGYRVMDOYADNAPVPOQATRAKQVIMVNO
CGMPREVTVAQGGQTNMTVTSGIKDGRKVVESIGIAGITGAKVTPKEWASSENQ
AAPSGVGTASEAPASEAK"
BASE COUNT      269 a      337 c      388 g      245 t
ORIGIN
Alignment Scores:
Pred. No.:      1,99e-110      Length:      1239
Score:          1989.00      Matches:      406
Percent Similarity: 99.03%      Conservative: 2
Best Local Similarity: 98.54%      Mismatches: 4
Query Match:    98.51%      Indels: 0
DB:             6      Gaps: 0
US-09-889-756a-2 (1-412) x AX391542 (1-1239)
QY      1 MetAlaPheTYrAlaPheLYeAlaMeTArGaLaAlaLeuAlaAlaValAlaLeu 20
Db      1 ATGGCTTTTATGCTTTTAAAGCGATGCGTCCGCGCTTGCGCCGCTTGCAITG 60
QY      21 ValLeuSerSerCYsGLyLYsGLyGLyAspAlaAlaGlnGLyGLyGlnProAlaGLyArg 40
Db      61 GTACTGCTGCTCTTCGCGTAAAGCGAGACCGGCGGCGGCGGCGGCGCTGCTGCTCG 120
QY      41 GluValaProAlaProValaValaGLyValaValaThyValaHisProGlnThyValaLeuThr 60
Db      121 GAAACCCCTCGCGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180
QY      61 ValGluLeuProGLyArGLeuGLuSerLeuArGThrAlaAspValaArgAlaGlnValaGLy 80
Db      181 GTCCAGTTGCGCGGCGGCTTTGAAATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
QY      81 GlyIleIleGlnLYsArGLeuPheGlnGLuGLySerTYrValaArgAlaGLyGlnProLeu 100
Db      241 GGCATCATCCAAAACGCTGCTTCCAGAAGCGATTATGTCGTCGCGGCGGCGGCGGCTG 300
QY      101 TyrGlnIleAspSerSerThyTYrGLuAlaAsnLeuGlnSerAlaArGAlaGlnLeuAla 120
Db      301 TATCAGATCGACAGTTCACCTTATGAAGCAGGCTCGGAAAGCGCGCGGCGCACTGGCA 360
QY      121 ThrAlaGlnAlaThrLeuAlaValaAlaAspAlaAspLeuAlaArGYrLYsProLeuVal 140
Db      361 ACGGCTCAGGCAAGCTTGCCTCAAGCGGATGCGGATTTGGCGCATTCACAGCTTTGATT 420
QY      141 AlaAlaGlnAlaValaSerArGLuGLyTYrAspAlaAlaValaThyAlaLYsArGLeAla 160
Db      421 GCGCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 480
QY      161 GluAlaGLyValaLYsAlaAlaGlnAlaAlaIleLYsSerAlaGLyIleAsnLeuAsnArg 180
Db      481 GAGGCGAGCGGCTTAAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
QY      181 SerArgIleThrAlaProIleSerGLyPheIleGLyGlnSerLYsValaSerGLyGlnThy 200
Db      541 TCGGCGATTACCGCGCGGATTTCCGCTTTATCGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
QY      201 LeuLeuAsnAlaGLyAspThyThyValaLeuAlaThyIleArgGlnThyAsnProMetTYr 220
Db      601 TTGCTGAACGCTGGCGATGCGACCGTACTGCGGCGACCAATCCGCAACCAATCCGATAT 660
QY      221 ValAsnValThyGlnSerAlaSerGLyValaMetLYsLeuArGLyGlnIleAlaGLyGLy 240
Db      661 GTGAACGTTACCCAGCTCGATCCGAAGTATGAAATTGGCGCGCTCAGATAGCGGAGGCG 720

```

```

QY      241 LysLeuLeuAlaAlaAspGLyValaIleAlaValaGLyIleLYsPheAspAspGLyThyVal 260
Db      721 AAATGCTGGCGCGGCGGATGGTGTATTCGGGTGCGCATCAAAATTGACGACGACGATT 780
QY      261 TyrProGLyLYsGLyArGLeuPheAlaAspProValaValaGlnGLySerThyGLyGln 280
Db      781 TACCTCAAAAAGCGCCCTGCTGTTGCGGATCCGCGCTCAACGAATCGACCGGCTCAG 840
QY      281 IleThrLeuArGLyAlaAlaValaProAspAspGlnIleLeuMetProGLyLeuTYrVal 300
Db      841 ATTACCTTGGCGCGCGCGGATCCGAACGATCAGATATCTTGATGCGCGGCTGTATGTG 900
QY      301 ArgValaLeuMetAspGlnValaAlaValaAspAsnAlaPheValaValaProGlnAlaVal 320
Db      901 CGCGTCTGATGACCAAGTGGCGGCTGATACCGCATTTGTTGCGCGGCGGCGGCGGCTA 960
QY      321 ThrArgGLyAlaLYsAspThyValaMetIleValaAsnAlaGlnGLyMetGLyProArg 340
Db      961 ACGGCGCGTCCGAAGAATACCGTATGATGATGATGATGATGATGATGATGATGATGATG 1020
QY      341 GluValaThyValaAlaGlnGlnGLyThyAsnThyIleValaThySerGLyLeuLYsAsp 360
Db      1021 GAGGTAAAGCTGTGGCAACAGCAGGCTACGAATGTGATGCTGCGGCTGGAAGCAC 1080
QY      361 GlyAspLYsValaValaGLyGLyIleSerIleAlaGLyIleThyArgLYsValaLYsVal 380
Db      1081 GGGGACAGAGTGTGTGGTGAAGCATCAGTATCCGCGTATACCGGCTGCGGCAAAAAGTA 1140
QY      381 ThrProLYsGLyThyAlaSerSerGLyAsnGlnAlaAlaAspProGlnSerGLyValaGln 400
Db      1141 ACGGCCAAAGATAGGCGGCTGCTGGAACCAAGCCCGCGCTCATCCGCGGCTTACG 1200
QY      401 ThrAlaSerGLyAlaLYsThyAlaSerGLyAlaGln 412
Db      1201 ACGGATCTGAAGCCAAACCTGCTTGAAGCGCAA 1236

```

RESULT 4
AE002521/c
LOCUS 14056 bp DNA linear BCT 25-MAY-2000
DEFINITION Neisseria meningitidis serogroup B strain MC58 section 163 of 206
ACCESSION AE002521 AE002098
VERSION AE002521.1 GI:7226962
KEYWORDS
SOURCE Neisseria meningitidis MC58
ORGANISM Neisseria meningitidis MC58
REFERENCE 1 (bases 1 to 14056)
AUTHORS Tettelin, H., Saunders, N.J., Heidelberg, J., Jeffries, A.C., Nelson, K.E., Eisen, J.A., Ketchum, K.A., Hood, D.W., Peden, J.F., Dodson, R.J., Nelson, W.C., Gwinn, M.L., Deboy, R., Peterson, J.D., Hickey, E.K., Haft, D.H., Salzberg, S.L., White, O., Fleischmann, R.D., Dougherty, B.A., Mason, T., Ciecko, A., Parksey, D.S., Blair, E., Citron, H., Clark, E.B., Cotton, M.D., Utterback, T.R., Khouri, H., Gin, H., Vamathevan, J., Gill, J., Scarlato, V., Maignani, V., Pizzi, M., Grandi, G., Sun, L., Smith, H.O., Fraser, C.M., Moxon, E.R., Rappelli, R. and Venter, J.C.
TITLE Complete genome sequence of Neisseria meningitidis serogroup B strain MC58
JOURNAL Science 287 (5459), 1809-1815 (2000)
MEDLINE 20175755
PUBMED 10710307
REFERENCE 2 (bases 1 to 14056)
AUTHORS Tettelin, H., Saunders, N.J., Heidelberg, J., Jeffries, A.C., Nelson, K.E., Eisen, J.A., Ketchum, K.A., Hood, D.W., Peden, J.F., Dodson, R.J., Nelson, W.C., Gwinn, M.L., Deboy, R., Peterson, J.D., Hickey, E.K., Haft, D.H., Salzberg, S.L., White, O., Fleischmann, R.D., Dougherty, B.A., Mason, T., Ciecko, A., Parksey, D.S., Blair, E., Citron, H., Clark, E.B., Cotton, M.D., Utterback, T.R., Khouri, H., Gin, H., Vamathevan, J., Gill, J., Scarlato, V., Maignani, V., Pizzi, M., Grandi, G., Sun, L., Smith, H.O., Fraser, C.M., Moxon, E.R.,

TITLE Rappunli, R. and Venter, J.C.
Direct Submission
JOURNAL Submitted (17-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
SOURCE Location/Qualifiers
1. .14056
/organism="Neisseria meningitidis MC58"
/mol_type="genomic DNA"
/strain="MC58"
/db_xref="taxon:122586"
/note="serogroup: B"
120. .1457
/gene="NMB1707"
120. .1457
/gene="NMB1707"
/note="similar to GB:AL009126 percent identity: 75.06;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="sodium- and chloride-dependent transporter"
/protein_id="AAF42055.1"
/db_xref="GI:7226963"

gene
CDS

gene
CDS
/translation="MSNNTSSSKIGFVLAAGSAGIGATWKPPTAGTNGAVFEL
LFLFTILALPQAEFYIGRTGKNAVDSFRVLRPEQMLWGRMGVAACTLLSF
YSVVGWLVNVVHSTGAVHAGTADFEALFGATISNPGSLSYQALFMTLVWVKG
ISDGIKXANRYLMFGLFILFALAIRSLTLPAGMGVSLKPNWSYFKADMTATAG
OAPFALSIGVAMITVASYLGKDDMPRSCHTMMNLLVSLAGLVFPVAFGE
PSGGGLIFIVLPVFMKMPRGVTLFVFMILVVPATLTSFMLETVIATSTRDOR
KRKHWTILGTAIFIGIPSLASFGWGEFVPGKTFIDLDVYISAVINIGALSLS
IFTAMIDKQSLVDKAGSTVPRAVLMLMTLRVLAIAIIVFINSIDL"
/complement (2297. .3091)
/gene="NMB1709"
/complement (2297. .3091)
/gene="NMB1709"
/note="similar to PID:2258280 percent identity: 98.49;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="chymidylate synthase"
/protein_id="AAF42056.1"
/db_xref="GI:7226964"

gene
CDS

gene
CDS
/translation="WKAYLDLMRHVLNDGTDKSDRTGTGTRSVFQYQMRFDLGKFPPL
LTTKMLHRSIHELMFELKGDITNKYKANNVSLWEMADENDLGVYGYQWRNP
APDGRHIDQIANVLEQIKNPDSRLIYSANNPALVDEMALPCHALFQFVAAGKLS
COLYORSADIFLGVPFNIAYSALLTMMMAVCGEAGEFVHTFGDAHLVYRHHFQOAL
OLERERPLPYMKINPEVKDLFSFKEDEFELEGDPHHIKAAVSV"
3475. .4809
/gene="NMB1710"
3475. .4809
/gene="NMB1710"
/note="similar to GP:146126 percent identity: 80.73;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="glutamate dehydrogenase, NADP-specific"
/protein_id="AAF42057.1"
/db_xref="GI:7226965"
/translation="WTDLTLLFANLKKORNPQOEPRHQAEEVFMSSLDPEFLANKRYTO
QSLERIVPERVWFRVTVWDQKQOVNNGYRVSASATGRTKGLRHPITVDLGV
LKPLAFEQVFKNALTTLPMGKGKGSDDPKGSDAEVWRCQAFMELVYHIDATD
VPADIGVGGREIIGYLPQYKKIRNEFSVLTGKLEWGSLLPEATGYCFFVAAQ
MLQTNDSFEGRVLISSGNVAOYAAEKALOLGKVLTVSDNGFVLPDPSGTEAO
LAALIEKEVEREVATAKEGLOYPFVKOKPQGVAAEIALPCATONLDEEAKKTL
ANGCYVABGANPESTLGAVERFITAAGLTXAPGRASNAGVATSGLENSQAIRLSST
REBVDRFLGIMQSIHESCLTKYKGVDTVNVNGANIAGFVKVADAMLQGF"
/complement (4944. .5723)
/gene="NMB1711"
/complement (4944. .5723)
/gene="NMB1711"
/note="similar to PID:455922 GB:U00096 SP:P06957
PID:1786303 PID:434009 percent identity: 53.25; identified
by sequence similarity; putative"
/codon_start=1

gene
CDS

gene
CDS
/transl_table=11
/product="transcriptional regulator, GntR family"
/protein_id="AAF42058.1"
/db_xref="GI:7226966"
/translation="MKLVPRQKISDOVLSYLERIAGVYAEAGCKIPPSVLAEEVY
SRPSVSALNVAVQVLEARQGGTYVSKFQDPLQSQMQLGKSNMEQVFPDS
CHLEGCAALAAERRTDADLKRTIGFWLEKEEESGEGSNHQSSEADVSFQTTADAAH
NLFSHLSGGLLMYLRQTRSSLIYNQEDPDPKMAQRVLYEAIISNRPEASBA
AKAHLNVYVASSIKDREYQSRNHADTLQNQDLKRVQDMVAV"
6154. .6453
/gene="NMB1712"
6154. .6453
/gene="NMB1712"
/note="similar to PID:882504 GB:U00096 SP:Q46839
PID:1789347 percent identity: 71.01; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="L-lactate permease-related protein"
/protein_id="AAF42059.1"
/db_xref="GI:7226967"
/translation="MORCFMPPHYCRBLEKRYRSGSIMEWTQNTAIGSLYLTAA
AALLPVFPFPAATVLEKGYQAGLTLLIALAAVAFGRMPGMVAVSLPPPH"
/complement (6839. .7804)
/gene="NMB1713"
/complement (6839. .7804)
/gene="NMB1713"
/note="similar to SP:P37247 PID:1197006 percent identity:
61.94; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="transposase, IS30 family"
/protein_id="AAF42060.1"
/db_xref="GI:7226968"

gene
CDS

gene
CDS
/translation="WSTYQLTQGERHYIQLNSHCYTEAKQNLNKKSTISREIRH
RTQGOQYSAKAGORRTIKQRRQPKLDSOLIHIDTIRKLSPEQVCAVLCXPH
RITLHSTIYRVLRQDSNGSLTMOHLRISCKEYRKGSTWTRGVNPGVLENRA
IVDOKSRITGMEADTVIGKQKSAALLTVRVTYRTIICLDLSKADETARAVALK
AHDORVHTITMDNGKEFYQHTKITKALKAEYVCPPYSHWEGKLENTVGLIRQYFPK
QTFPNSIDREIRVQDELNHRPRKTLGYTPSVLFLNLQPLIH"
/complement (7914. .9317)
/gene="NMB1714"
/complement (7914. .9317)
/gene="NMB1714"
/note="similar to GB:L23839 PID:404773 PID:833173 percent
identity: 64.15; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="multidrug efflux pump channel protein"
/protein_id="AAF42061.1"
/db_xref="GI:7226969"
/translation="WDTTEKTTITSVAAAFALSACTMIPYEOPKVEVATFRKDTAD
SGIRAVDLGMHDFADPRLOKLDIALERTSLRTAVLNSIYRKQWIERNNLLPL
AAVANDRSQSLSGGVSSYKVLGAASVELDFGVRSSEALQGYFAVARNRA
AHLSLIATVAKAVENERYAEAMSLAQRVLTETKLSLEYLAKAVIAVALLROE
ALIESAKADVAHARSROARNALATINQIPEDDPAGPLDQFVEVXLPGISSE
VLDRPPIRAEAHALKQANMIGAAAPAFPSIRLCTGTGSAEIGCLFKSGGWS
FASITPIPTTWGNKANDVAKLRQOVQVAVESAVQSFQVYANALARBOLDKAY
DALSKGRARKEALRLVGLRYKGVSGALDLDLAERSVYABGAALSAQLTRANELAD
LYKLAGGLRQEDTQDK"
/complement (9372. .12575)
/gene="NMB1715"
/complement (9372. .12575)
/gene="NMB1715"
/note="similar to PID:1408202 percent identity: 98.59;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="multiple transferable resistance system protein
MtrD"
/protein_id="AAF42062.1"
/db_xref="GI:7226970"

ORIGIN

Alignment Scores:

Pred. No.: 3,18e-107 Length: 349980
 Score: 1989.00 Matches: 406
 Percent Similarity: 99.03% Conservative: 2
 Best Local Similarity: 98.54% Mismatches: 4
 Query Match: 98.51% Indels: 0
 DB: 6 Gaps: 0

US-09-889-756a-2 (1-412) x AX044033 (1-349980)

```

Qy      1 MetAlaPheTyrAlaPheLysAlaMetArgAlaAlaAlaLysAlaAlaValAlaLeu 20
Db      296382 ATGGCTTTTATGCTTTTAAAGCCGATGCGTGGCGCGCTGGCTGGCGCGCTGGCATTTG 296323

Qy      21 ValLeuSerSerCyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 40
Db      296322 GTACTGTGCTTGGGTTAAAGCGGAGACGCGGCGGAGCGGCGGCGGCGGCGGCTGCTGG 296263

Qy      41 GluAlaProAlaProAlaValAlaGlyValAlaThrValHisProGlnThrValAlaLeuThr 60
Db      296262 GAAGCCCTCGCGCGCGCTGCGGTGCGTGAACCGTCCATCCGCAAAACCGTCCATTGACC 296203

Qy      61 ValGluLeuProGlyArgLeuGluSerLeuArgThrAlaAspValArgAlaGlnValGly 80
Db      296202 GTCGAGTTGCCGCGGCGCTTGGAAATCGCTGCGTACCGCGCATGTCGCGCCCAAGTCGCGC 296143

Qy      81 GlyIleIleGlnLysArgLeuPheGlnGlnGlySerTyrValArgAlaGlyGlnProLeu 100
Db      296142 GGCATCATCCAAAACGCTGTTCAGAAAGCGACTTATGTCGCGCGGAGCAGCGCTG 296083

Qy      101 TyrGlnIleAspSerSerThrTyrGlnAlaAsnLeuGlnSerAlaArgAlaGlnLeuAla 120
Db      296082 TATCGATCGACAGTTCCACTTATACAGCAGCTCTGGAAGCGCGCGCGCACTGCGCA 296023

Qy      121 ThrAlaGlnAlaThrLeuAlaLysAlaAspLeuAlaArgTyrLysProLeuVal 140
Db      296022 ACGGCTGAGCAACGCTTCCAAACGCGATGGGATTTGGCGCGATACAAAGCTTTGGTT 295963

Qy      141 AlaAlaGluAlaValSerArgGlnGlyTyrAspAlaAlaValThrAlaLysArgSerAla 160
Db      295962 GCCCGCGAAGCGCTGACGCGGAGAGAAATACGATGCTGGGTAAACGCGCAAAAGCTTCTCC 295903

Qy      161 GluAlaGlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsnArg 180
Db      295902 GAGGACGCGCTTAAAGCGCGCGGCGCAATCAATCCGCGCATCAGCTGAACCGT 295843

Qy      181 SerArgIleThrAlaProLysSerGlyPheIleGlyGlnSerLysValSerGlnGlyThr 200
Db      295842 TCGGCGATTACCGCGCGCATTTCCGCTTTATCGGTCAGTCCAAAGTTTCCGAAGGTACG 295783

Qy      201 LeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAspProMetTyr 220
Db      295782 TTGCGAAGCGCTGGGATGCGACCGCTACTGCGACATCCGCCCAACCAATCCGATGTAT 295723

Qy      221 ValAsnValThrGlnSerAlaSerGlyValMetLysLeuArgArgGlnIleAlaGlnGly 240
Db      295722 GTGAACGTTACCCAGCTGCTGATCCGAAGTGATGAATTTGCCCGCTCAATGCCGAAGC 295663

Qy      241 LysLeuLeuAlaAlaAspGlyValIleAlaValGlyIleLysPheAspAspGlyThrVal 260
Db      295662 AAACGCTGGCGCGCGATGCTGTGATTCGCGTCCGATCAATTTGACGACGCGCACGTT 295603

Qy      261 TyrProGlnLysGlyArgLeuLeuPheAlaAspProValValAsnGlnSerThrGlyGln 280
Db      295602 TACCTCGAAAAAGCGCGCTGCTGTTCGCCATCCGCGCGCAAGAAATCGACCGCGTAG 295543

Qy      281 IleThrLeuArgAlaAlaValAlaProAsnAspGlnAsnIleLeuMetProGlyLeuTyrVal 300
Db      295542 ATTACCTCGCGCGCGCGGATCGACCAATCAATCAATCTTGATGCTCCGCTGTATATG 295483

Qy      301 ArgValLeuMetAspGlnValAlaValAspAsnAlaPheValValProGlnGlnAlaVal 320

```

```

Db      295482 CGCGTCTGATGACCAAGTGCGCGTGATACGCAATTGTGTGCGCGACGCGGTA 295423

Qy      321 ThrArgGlyAlaLysAspThrValMetIleValAsnAlaGlnGlyGlyMetGluProArg 340
Db      295422 ACGCGCGGTGGAAAGATACCGATGATGTGATGTGATCCCAAGCGGTTATGAAACCCCGC 295363

Qy      341 GluValThrValAlaGlnGlnGlnGlyThrAsnTrpIleValThrSerGlyLeuLysAsp 360
Db      295362 GAGGTAAAGGTGGGACCAAGCAGCGGTACGAATTTGATGTGATGCTGGGTCTGAAGAC 295303

Qy      361 GlyAspLysValValAlaGlnGlyLysSerIleAlaGlyIleThrGlyAlaLysVal 380
Db      295302 GCGGACAAAGATGGGGGTGTGTGGAAGCATCATGCGCGCTTAAACGGGTGCGAAAGATA 295243

Qy      381 ThrProLysGluTrpAlaSerSerGluAsnGlnAlaAlaAlaProGlnSerGlyValGln 400
Db      295242 ACGCCCAAGATGGGGGTGTGTGGAAGCATCATGCGCGCTTAAACGGGTGCGAAAGATA 295183

Qy      401 ThrAlaSerGluAlaLysThrAlaSerGluAlaGlu 412
Db      295182 ACGGCATCTGAGCGCAAAACCTGCTTGAAGCGAAA 295147

RESULT 6
NGU14993
LOCUS
DEFINITION
Neisseria gonorrhoeae membrane fusion protein (mtrc) gene, complete cds.
ACCESSION
U14993
VERSION
U14993.1 GI:550459
KEYWORDS
Neisseria gonorrhoeae
ORGANISM
Neisseria gonorrhoeae
Bacteria: Proteobacteria; Betaproteobacteria; Neisseriales;
Neisseriaceae; Neisseria.
REFERENCE
1 (sites)
Hagman,K.E., Pan,W., Spratt,B.G., Balhazar,J.T., Judd,R.C. and
Shafer,W.M.
Resistance of Neisseria gonorrhoeae to antimicrobial hydrophobic
agents is modulated by the mtrRDE efflux system
Microbiology 141 (Pt 3), 611-622 (1995)
JOURNAL
MEDLINE
95227358
PUBMED
7711899
REFERENCE
2 (bases 1 to 1500)
Pan,W. and Spratt,B.G.
Regulation of the permeability of the gonococcal cell envelope by
the mtr system
Mol. Microbiol. 11 (4), 769-775 (1994)
JOURNAL
MEDLINE
94254732
PUBMED
8196548
REFERENCE
3 (bases 1 to 1500)
Shafer,W.M.
Direct Submission
Submitted (21-SEP-1994) William M. Shafer, Microbiology/Immunology,
Emory University, 1510 Clifton Road, Atlanta, GA 30322, USA
FEATURES
source
1..1500
/organism="Neisseria gonorrhoeae"
/mol_type="genomic DNA"
/strain="FAL9"
/db_xref="taxon:485"
45..57
/function="protein binding target"
241..244
RBS
251..1489
/gene="mtrC"
251..1489
/gene="mtrC"
251..1489
/note="mtrC"
/citation=[1]
/codon_start=1
/transl_table=1
/product="membrane fusion protein"

```



```

gene
1.1235
/gene="mtc"
/pseudo
1.1232
/gene="mtc"
/note="membrane fusion protein MtrC; translation
interrupted by frameshift error"
/pseudo
/codon_start=1
/transl_table=11

BASE COUNT      270 a    332 c    387 g    246 t
ORIGIN
Alignment Scores:
Pred. No.:      1,27e-105      Length:      1235
Score:          1908.50      Matches:      391
Percent Similarity: 97.09%      Conservative: 9
Best Local Similarity: 94.90%      Mismatches: 11
Query Match:      94.53%      Indels:      2
DB:                1      Gaps:        1

US-09-889-756a-2 (1-412) x AF037041 (1-1235)

Qy      1 MetAlaPheTyrAlaPheLysAlaMetArgAlaAlaAlaLeuAlaAlaValAlaLeu 20
Db      1 ATGGCTTTTATGCTTCTTAAGCGCATGCGCGCGCGCTGGCTGGCGCGCTTGCAATTG 60
Qy      21 ValLeuSerSerCysGlyLysGlyGlyAspAlaAlaGlnGlyGlyGlnProAlaGlyArg 40
Db      61 GCACGTGCTGCTTGGGTAAAGCGGAGACCGGCGCGGCGGCGGACCTGCGGCTGG 120
Qy      41 GluAlaProAlaProValAlaGlyValAlaThrValHisProGlnThrValAlaLeuThr 60
Db      121 GAAGCCCCCGCGCGCGCTGCTGCGCTGCTAACCGTCCATCCGCAAAACCGTGCATTGAC 180
Qy      61 ValGluLeuProGlyArgLeuGlnGlnSerLeuArgThrAlaAspValAlaGlnValGly 80
Db      181 GTCCAGTGGCCGGGCGCTTGGAAATCCGTGCGTACCGCGCATGTCGCGCCCAAGTCGCG 240
Qy      81 GlyIleIleGlnLysArgLeuPheGlnGlnGlnSerTyrValAlaArgAlaGlyGlnProLeu 100
Db      241 GGCAATCATCCAAAGCGCTGTTCACAAAGCGGATTATGTCGCGCGCGGCGAGCGCTG 300
Qy      101 TyrGlnIleAspSerSerThrTyrGlnAlaLeuGlnSerAlaArgAlaGlnLeuAla 120
Db      301 TATCAGATCGACAGTCCACTTATGACAGCAGTCTCGAAG-----CGCGCCCACTGSCA 356
Qy      121 ThrAlaGlnAlaThrLeuAlaLysAlaAspAlaAspLeuAlaArgTyrLysProLeuAl 140
Db      357 ACGGCACAGGCAACGCTTGCACAAAGCGGATGCGGCTTGGCGCTTACAAACCGCTGCTT 416
Qy      141 AlaAlaGlnAlaValSerArgGlnGlnLysTrpAspAlaAlaValThrAlaLysArgSerAl 160
Db      417 TCCGCCGATGCGCATGTAACCAAGATGATGTCGTGTAACCGGCAAAACGTTCTGCC 476
Qy      161 GluAlaGlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsnArg 180
Db      477 GAGCGGGCGCTTAAAGCGGCGGCGGCGGCGGATCCGCGCGGATCAATCTGAACCT 536
Qy      181 SerArgIleThrAlaProIleSerGlyPheIleGlnGlnSerLysValSerGlnLysThr 200
Db      537 TCGGCATTTACCGCGCGCATTTCCGGCTTATCGGTCAAGTTCGGAAGATGACG 596
Qy      201 LeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAsnProMetLys 220
Db      597 CTGTTGAATGCGGGGATACAACTGTTTACCAACATCCGCAAAACCAATCCGATGTAT 656
Qy      221 ValAsnValThrGlnSerAlaSerGlnValMetLysLeuArgArgGlnIleAlaGlnGly 240
Db      657 GTGAACGTTACCCAGTCTGCATCCGAAGTGAAGTAACCTCCGCGGCAATAGCCAAAGC 716
Qy      241 LysLeuLeuAlaAlaAspGlyValIleAlaValGlyIleLysPheAspAspGlyThrVal 260

```

```

Db      717 AAGCTGTCGCGCGGATGCGGATTCGATCGGCGCATCAATTGACGACGATCGGTT 776
Qy      261 TyrProGlnLysGlyArgLeuLeuPheAlaAspProValAlaAsnGlnSerThrGln 280
Db      777 TATCTTAAAAAGGCGCGCTGCTGTTTGTGATCCGACCGTTGACGATGACCGGTCAG 836
Qy      281 IleThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetProGlyLeuTyrVal 300
Db      837 ATTACCTTGGCGCGCGCGCTATCGAACGATCAGATATCTTGATCCCGCGCTGTATGTG 896
Qy      301 ArgValLeuMetAspGlnValAlaValAspAsnAlaPheValValProGlnGlnAlaVal 320
Db      897 CGTGTCTGATGATCATCGATGCGCGCGGATTAATGATTCATCGTCCACAGCGCGGTA 956
Qy      321 ThrArgGlyAlaLysAspThrValMetIleValAsnAlaGlnGlnGlyMetGluProArg 340
Db      957 ACGGCGCGTTCGAAAGATACCGTATATCTGATATCCACAGCGGCTATGAAACCCCGC 1016
Qy      341 GluValThrValAlaGlnGlnGlnGlyThrAsnTrpIleValThrSerGlyLeuLysAsp 360
Db      1017 GAGGTACCGTCCGCGCAACAGCAGGATGACAAATTGATGTTACGTGCGGCTGAAAGAC 1076
Qy      361 GlyAspLysValValAlaGlnGlyIleSerIleAlaGlyIleThrGlyAlaLysLysVal 380
Db      1077 GGGGACAAAGTGTGTGGAAGGCATCATGATCCCGGTATGACGGGCGGAAAAAGGTA 1136
Qy      381 ThrProLysGlnTrpAlaSerSerGluAsnGlnAlaAlaAlaProGlnSerGlyValGln 400
Db      1137 ACGCTTAAAGATGCGCGCGCTCTGAAATCATACCTCCCGCTTCAAGCGCGCTTCAG 1196
Qy      401 ThrAlaSerGlnAlaLysThrAlaSerGlnAlaGln 412
Db      1197 ACGGCATCTGAAGCAACCTGCTTGAAGCGGAA 1232

RESULT 8
NGMTREC/c
LOCUS      2127 bp      DNA      linear      BCT 01-OCT-1994
DEFINITION N.gonorrhoeae mtrr and mtrc (partial) genes.
ACCESSION  Z25796
VERSION     Z25796.1 GI:438190
KEYWORDS   mtrc gene; mtrr gene.
SOURCE      Neisseria gonorrhoeae
ORGANISM   Neisseria gonorrhoeae
            Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
            Neisseriaceae; Neisseria.

REFERENCE   1
AUTHORS    Pan, W. and Spratt, B.G.
TITLE      Regulation of the permeability of the gonococcal cell envelope by
            the mtr system
JOURNAL     Mol. Microbiol. 11 (4), 769-775 (1994)
MEDLINE     94254732
PUBMED     8196548
REFERENCE   2 (bases 1 to 2127)
AUTHORS    Spratt, B.G.
TITLE      Direct Submission
JOURNAL     Submitted (25-ARG-1993) Spratt B. G., University of Sussex, School
            of Biology, Falmer, Brighton, East Sussex, UK, BN1 9QG
FEATURES
            Location/Qualifiers
                1..2127
                    /organism="Neisseria gonorrhoeae"
                    /mol_type="genomic DNA"
                    /strain="CH95 (multiply antibiotic-resistant strain)"
                    /db_xref="taxon:485"
                    complement(1..814)
                    /gene="mtrc"
                    complement(<1..814)
                    /gene="mtrc"
                    /functions="Regulates permeability of cell envelope"
                    /note="The level of expression of the mtrc protein appears
                    to determine the susceptibility to hydrophobic compounds.
                    MtrC is homologous to EnvC and AcrA of Escherichia coli."
                    /citation=11
                    /codon_start=1

```

```

/evidence=experimental
/transl_table=11
/product="Mtrc protein"
/protein_id="CAAB1046.1"
/db_xref="GI:438192"
/db_xref="SPTREMBL:O51007"
/translation="WAFYASKRAAAALAAVALATSSCGKRDAAOCGAPAGAPAPVGVGHGHPOTVALTVLPGRLESLRADVAOVGGIIGQRLEDEGSYVRAGPVLQIDSDTYEAGLSARQAQLTAQNTTAKADADLARKEPLVASIAISKOETDAVTKRSEFAGVAAQAAIKASGINLRSRITAPISGFGOSKVSBSGLINAGDTLVATIKQTPNMVVANTQSASEVMKLRRQIAEGKLAADGAIAVGHKPDGTVEPEKKLLPAD"
1064..1696
/gene="mtrr"
/locus_tag="MTRR_1064"
/feature_key="misc_feature"
/function="putative repressor of MtrC gene"
/note="The mtrr gene encodes a putative repressor that controls the permeability of the N. gonorrhoeae cell envelope. A homologous gene is found divergently transcribed from both E. coli envC and acra."
/citation=[1]
/codon_start=1
/evidence=experimental
/transl_table=11
/product="Mtrr protein"
/protein_id="CAAB1045.1"
/db_xref="GI:438191"
/db_xref="SPTREMBL:O59608"
/translation="WRKTTEALKTKHEMLMALETFRYKGARTSIINELAQAGVTRGALVWHFNKEDFDLAFORICDIENTIADADAEGBSWTPRHLLHFFPERLOSNDLYYKFENILFLKCHTEONAAVIALARKHAIWKRTAVLTAVENODLADDLKEN TAVFIKSTLDGLIWRWSSGESFDLGKTAIRTIIGIMDNLENHPCRK"
1715..1725
/product="uptake sequence"
/note="An inverted repeat of the gonococcal uptake sequence is found downstream of mtrr. The inverted uptake sequences may act as a transcription terminator - a common feature of gonococcal genes."
/citation=[1]
/function="DNA uptake in transformation"
/complement(1737..1747)
/product="uptake sequence (9/10 match to consensus)"
/note="Part of an inverted repeat composed of two uptake sequences (see feature d)"
/citation=[1]
/function="DNA uptake in transformation"
2102..2125
/product="Repeated sequence in Neisseria genome"
/note="This sequence is found downstream of a number of N. gonorrhoeae and N. meningitidis genes - including pilin genes, transferrin-binding protein genes, opa genes and within IS1106."
/citation=[1]
/function="Unknown"
/phenotype="Unknown"
BASE COUNT      550 a      573 c      481 g      523 t
ORIGIN
Alignment Scores:
Pred. No.:       1,2e-67          Length:         2127
Score:           1278.00          Matches:         262
Percent Similarity: 98.15%        Conservative:     4
Best Local Similarity: 96.68%      Mismatches:      5
Query Match:      63.30%          Indels:          0
DB:               1              Gaps:            0

US-09-889-756A-2 (1-412) x NCMTTRC (1-2127)
Oy 1 MetAlaphETyrAlaPhelYalawerArGa1aa1aalenu1aa1aa1ava1aleu 20
    |||||
Db 814 ATGGCTTTATTACGCTTCAGGCAGCATGCTGCCGCCGCGCTTGCGCCGCTTGCAATTG 755
    |||||
Oy 21 ValIeuSerCysGlyIlyGlyGlYAspa1laa1aIng1yglInPrCoAlag1Arg 40
```

Db	Accession	Gene	Protein	Length	
Db	754	GCACGTGCTGCTTCCGGTAAGGACAGAGACCCGGCGCAGGGCCGGCAGCCCTCGCGGCTGG		695	
Qy	41	Glua1aProalAProVal1aGlyVal1aThrValHisProGlnThrVal1aLeuThr		60	
Db	694	GAAACCCCGCGCCGCTGCTCGGCGGTGAACCGTCCATCCGCAAAACCGTCGATTCAGC		635	
Qy	61	Val1aGluLeuProGly1aArgLeuGlnSerLeuArgThr1aAspVal1aArg1aGlnVal1aGly		80	
Db	634	GTCAGAGTTCCGGGGCGGTGGGAATCGCTGGTACCGCCGATGTCGCGGCCCAAGTCGCG		575	
Qy	81	Gly1aLeilegln1aArgLeuPheGlnGln1aSer1aVal1aArg1aGly1aProLeu		100	
Db	574	GGCATTCATCCAAAACCGCTCTTCCAAAGAGGCGATTATGTCGCCGCCGGCAGCCGCTG		515	
Qy	101	TyGln1aLeaSerSer1aThr1aGln1aAsnLeuGlnSer1aArg1aGln1aLeu1a		120	
Db	514	TATCAGATCGACAGTTCACATTATGAAGCAGAGTCTGGAAAGCGCGCGCCCACTGGCA		455	
Qy	121	Thr1aGln1aThr1aLeu1aVal1aAsp1aAsp1aLeu1aArg1aGly1aProLeuVal		140	
Db	454	ACGGCACAGGCAAGCGCTTGGCCAAAGCGGATGCGGATTTGGCGCTTCMAACCGCTGGTT		395	
Qy	141	Ala1aGln1aVal1aSer1aArg1aGln1aThr1aAsp1aVal1aThr1aVal1aArg1aSer1a		160	
Db	394	TCCGCCGATGCCATCAGTAACAAGAGTACGATCTCGGTAAGCGGCAACGTTCTGCC		335	
Qy	161	Glu1aGly1aVal1aVal1aAla1aGln1aAla1aLeu1aSer1aGly1aLeu1aAsn1aArg		180	
Db	334	GAGCGGGCGGCTTAAGGGCGGCGAGCGCGCATCAAAATCCGCGGCATCAATCTGAACGCT		275	
Qy	181	Ser1aGly1aThr1aPro1aLeu1aSer1aGly1aPhe1aGly1aGln1aSer1aVal1aSer1aGly1aThr		200	
Db	274	TCCGCGCATTCGCGCGCGATTTCCGGCTTTATCCGTACATCCAAAGTTTCCGAAGTACG		215	
Qy	201	Leu1aAsn1aGly1aAsp1aThr1aThr1aVal1aLeu1aThr1aLeu1aGln1aThr1aAsn1aPro1aMet1aThr		220	
Db	214	CTGTTGAATGGGGCGCATCAACTGTTTACCCACCATCCGCGCAAAACCAATCCGATGTAT		155	
Qy	221	Val1aAsn1aThr1aGln1aSer1aAsp1aVal1aMet1aLeu1aArg1aArg1aGln1aLeu1aGly1a		240	
Db	154	GTCGACCTTACCCAGCTGTCATCCGAAGTATGAATCTCGCGCGCGCATGACCGGAAGGC		95	
Qy	241	Leu1aLeu1aAla1aAsp1aGly1aVal1aAla1aVal1aGly1aLeu1aPhe1aAsp1aGly1aThr1aVal		260	
Db	94	AAGCTGCTGGCGGGCGGATGGTGCATTCGCGTCGCGATCAAAATTTGACGACGTAACGTT		35	
Qy	261	Tyr1aProGln1aGly1aArg1aLeu1aPhe1aAsp		271	
Db	34	TATCTGAAAAAAGCCCGCTGCTGTTCGTAT		2	
RESULT 9	AL646057/c	215050 bp	DNA	linear	BCT 02-SEP-2002
LOCUS	AL646057	215050 bp	DNA	linear	BCT 02-SEP-2002
DEFINITION	Ralstonia solanacearum GM11000 chromosome, complete sequence;				
ACCESSION	AL646057				
VERSION	AL646057.1				
KEYWORDS	GI:17427008				
SOURCE	Ralstonia solanacearum				
ORGANISM	Ralstonia solanacearum				
REFERENCE	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;				
AUTHORS	Burkholderiaceae; Ralstonia.				
1	Salanoubat M., Genin S., Artiguenave F., Gouzy J., Manganot S.,				
	Atlat M., Billault A., Brotier P., Cenus J.C., Catolico L.,				
	Chandler C., Choisme N., Cladel-Renaud C., Cunnac S., Demange N.,				
	Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,				
	Signier P., Thebaud P., Whalen M., Wincker P., Levy M.,				
	Weissenbach J. and Boucher C.A.				
TITLE	Genome sequence of the plant pathogen Ralstonia solanacearum				
JOURNAL	Nature 415 (6871), 497-502 (2002)				
MEDLINE	21681879				

TITLE Stoneking, T., Nhan, M., Waterston, R. and Wilson, R. K.
Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2

JOURNAL Nature 413 (6858), 852-856 (2001)

MEDLINE 21534948

PubMed 11677609

REFERENCE 2 (bases 1 to 20382)

AUTHORS

CONSRITM

TITLE The *Salmonella typhimurium* Genome Sequencing Project

JOURNAL Direct Submission

Submitted (29-MAR-2001) Genome Sequencing Center, Department of Genetics, Washington University School of Medicine, 4444 Forest Park Boulevard, St. Louis, MO 63108, USA

COMMENT Supported by NIH grant 5U 01 AI43283

COMMENT

Coding sequences below are predicted from manually evaluated computer analyses, using similarity information and the programs; GLIMMER; <http://www.tigr.org/softlab/glimmer/glimmer.html> and Genemark; <http://opal.biology.gatech.edu/Genemark/>

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto Encyclopedia of Genes and Genomes; <http://www.genome.ad.jp/kegg/>, and Pedro Romero and Peter Karp at ECoCyc; <http://ecocyc.PangeaSystems.com/ecocyc/>

The analyses of ribosome binding sites and promoter binding sites were kindly provided by Heladiada Salgado, Julio Collado-Vides and ReguonDB; http://knich.cifn.unam.mx:8850/db/regulondb_intro.frameset

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternative chemistries or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

location/Qualifiers

1..20382

/organism="Salmonella typhimurium LT2"

/mol_type="genomic DNA"

/strain="LT2; SGGC 1412; ATCC 700720"

/db_xref="ATCC:700720"

/db_xref="taxon:99287"

/note="LT2"

195..729

/gene="accB"

/note="synonym: STM3379"

195..203

/gene="accB"

/note="putative -35 signal for accB; RegulonDB:STMLH004589"

221..229

/gene="accB"

/note="putative -10 signal for accB; RegulonDB:STMLH004589"

246..251

/note="accB"

/note="putative RBS for accB; RegulonDB:STMS1H003389"

259..729

/gene="accB"

/EC_number="6.4.1.2"

/note="carrier of biotin; similar to E. coli acetylCoA carboxylase, BCCP subunit; carrier of biotin (AAC76287.1); Blastp hit to AAC76287.1 (156 aa), 92% identity in aa 1 - 156"

/codon_start=1

/transl_table=1

/product="acetyl CoA carboxylase, BCCP subunit"

/protein_id="AA122248.1"

/db_xref="GI:16421940"

/translation="MDIRIKKILIELVESISLEISSEGESVRISFTTANAGFPVM QQAAAPMGOOPALSNAAVAPATPAMEAPAAAEISGHVRSFPMVGTFYRTPSPAKAF IEVQKVNVDLCTIVEAMKMNQIEADKAGTVAIVESGQVPEFDEPLVIE"

gene 731..2089

/gene="accC"

/note="synonym: STM3380"

RBS 731..736

/gene="accC"

/note="putative RBS for accC; RegulonDB:STMS1H003390"

CDS 740..2089

/gene="accC"

/EC_number="6.3.4.14"

/note="similar to E. coli acetyl CoA carboxylase, biotin carboxylase subunit (AAC76288.1); Blastp hit to AAC76288.1 (449 aa), 96% identity in aa 1 - 449"

/codon_start=1

/transl_table=1

/product="acetyl CoA carboxylase"

/protein_id="AA122249.1"

/db_xref="GI:16421941"

/translation="MDKIVIANRGEIALRILRACKELGIKTVAVHSSADRLKHVLL ADEIVICGPAPSVKSYLNIIPALISAETGAVAIHPQYGLSENANPAEOVERSGFLF IGPADITIRLMGDKVSAITAMKRGVPTVPGSDGLDDMKANRAHAKRIGYPIVIA SGGGGGGRMVRVSDAELAOISMTKAAKASNDVMNKEITENPHRIEIOVLADG OGNAIYLAERDCSMORRHOKVVEAPAPGITPLRRIYIGRCARACADICVYGATPE FLRENGEFTYENMTRIQEHPTMTITGDILKEQRIAGOPLSITQEBVYRGHA VECRINAEEDNTPLSPGKITRPHAPGCFVNRSHIYAGITVPYVDSMIGKLTG ENRDVAIARMKNAQLDILIDGIRKTNIDLOTRINMDEHFQHGNTIHYLEKKGLOEK"

2186..2440

/gene="yhdT"

/note="synonym: STM3381"

2186..2191

/gene="yhdT"

/note="putative RBS for yhdT; RegulonDB:STMS1H003391"

2198..2440

/gene="yhdT"

/note="similar to E. coli orf, hypothetical protein (AAC76289.1); Blastp hit to AAC76289.1 (80 aa), 82% identity in aa 1 - 80"

/codon_start=1

/transl_table=1

/product="putative inner membrane protein"

/protein_id="AA122250.1"

/db_xref="GI:16421942"

/translation="MDARFVOAHKEARMAWLTLCYAAWLVAAVYLPDGSPTGLDPH WFEWACLTPLVPLVILCWMAVVKFIYRDISLEDDAA"

2417..3881

/gene="panF"

/note="synonym: STM3382"

2417..2422

/gene="yhdT"

/note="putative RBS for panF; RegulonDB:STMS1H003392"

2430..3881

/gene="panF"

/note="similar to E. coli sodium/pantothenate symporter (AAC76290.1); Blastp hit to AAC76290.1 (485 aa), 93% identity in aa 3 - 485"

/codon_start=1

/transl_table=1

/product="SSS family sodium/pantothenate symporter"

/protein_id="AA122251.1"

/db_xref="GI:16421943"

/translation="MDLEVILPVAVYLVVFGVSIYAMRRKTAGTFLNXYFLGSRSGG GIVAMTLATVYISASSFIGPGAAVYGVGWLAMTQLPAWLSGLIGKFAIILA RRYNAVTLNMLFARVOSRLVWLWASISLVAFIGMTVQFISGARLLETGAGIPIYT GLIFGVISMLYTAFFGFRASVNDTLOGVMVGGITVLVGVYHAAGGSGQAVDTLH ALDPKIVTPGAGDDITISPARMTSPWLVGVGIGLPHRTAVRCISYKDSKAVHGIIIG TIVAVITMFGMHLAAGLGRAVLDITVYPTDITVITWKKVLPFAAGIFLAAPAAIIS TINQQLQSSATITIKDLYLNLRPDQMNELRLKMSAATLLGLALLLAKMPPEMT IVMNLAFGLAEAVFLMPLVLGLGYWEANAGALSAWIVGGVYVALLATENTIOYLGPH PIVPALILSLAFLIGNRFSSASQATVISTDR"

3893..4774

/gene="prnA"

/note="synonym: STM3383"

3893..4774

/gene="prnA"

```

/EC_number="2.1.1."
/note="similar to E. coli methylase for 50S ribosomal
subunit protein L11 (AAC76291.1); Blastp hit to AAC76291.1
(293 aa), 95% identity in aa 1 - 292"
/codon_start=1
/transl_table=11
/product="methylation of 50S ribosomal subunit protein
L11"
/protein_id="AAL22252.1"
/db_xref="GI:16421944"
/translation="MPPITQKLTNTGANAEEISDALMEAGAVSTFDTHDTPYEPL
FGEIRLMGDDTVIGLFDPAETDKQVNAILEGHPULGAGFAKIBQSDKMEREMDN
FHPERFGERLWI CPSMRDIPDENAVNMLDGLAFGTHPTTSLCQWMLDGLDNK
TVIDFGCGSGILATLAKLGAAKAIGIDIDPOAIQASRDNAENGVSDRLYLPKQ
PEAKKADVVVANIILAGPLRELAPLISVLPVGGILGSGILASQASVCDAVALFTL
DPVYKEKEMCKITGRKK"
gene
5326.6400
/ gene="yhdG"
/ note="synonym: STM3384"
5326.5334
/ gene="yhdG"
/ note="putative -35 signal for yhdG;
RegulonDB:STM1H004590"
5352.5368
/ gene="yhdG"
/ note="putative binding site for FIS, RegulonDB:
STM1H000148"
5353.5361
/ bound_molety="FIS"
5353.5361
/ gene="yhdG"
/ note="putative -10 signal for yhdG;
RegulonDB:STM1H004590"
5417.5422
/ gene="yhdG"
/ note="putative RBS for yhdG; RegulonDB:STM1H003393"
5435.6400
/ gene="yhdG"
/ note="hypothetical 35.9 kDa protein in pmrA-fis
intergenic region (ORF1). (SW:YHDG_SALTY); possibly
US-09-889-756a-2 (1-412) x AE008856 (1-20382)
Alignment Scores:
Pred. No.: 1.95e-36 Length: 20382
Score: 777.00 Matches: 170
Percent Similarity: 60.25% Conservative: 71
Best Local Similarity: 42.50% Mismatches: 121
Query Match: 38.48% Gaps: 38
DB: 1 Indels: 5
US-09-889-756a-2 (1-412) x AE008856 (1-20382)
QY 19 AlaLeuValIeuSerSerCyGlyLyGlyAspAla---AlaInglyGlyInPro 37
DB 11319 GCTGCGCTGCTGCGGCTTGAATGACCGAGAGATCCCGAGCTCAGCGGAGCCG 11378
QY 38 AlaGlyArgGluAlaProAlaProValValGlyValThrValHis-----Pro 54
DB 11379 CAA-----GTCAACCGTCATGTGCTGGAACA 11405
QY 55 GlnThrValAlaLeuThrValGluLeuProGlyArgLeuGluSerLeuArgThrAlaAsp 74
DB 11406 GCGCGCGTACCGCTTAAAGCACTTCCGAGTCGTCGCCGATTCGATTCGCGAG 11465
QY 75 ValArgLagInValGlyGlyIleLeuGlnLysArgLeuPheGlnGlnGlySerTyrVal 94
DB 11466 GTTCGCCCCCGAGGTGACCGGAGATGCTTAAAGAACTTCAACCGAAGAGGATGTA 11525
QY 95 ArgAlaGlyInProLeuTyrGlnIleAspSerSerThrTyrGlyAlaAsnLeuGluSer 114
DB 11526 GAGCGCGGAGGTGCTCTATCATGATCGATCCGCACTTATCGATTAAGACAGC 11585
QY 115 AlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaLysAlaAspAlaAspLeuAla 134
DB 11586 GCTAAAGCGCAACTTGTAAAGCGAAGCGGCTGCGCTATCGCGCACCTGACGCTCAA 11645

```

```

QY 135 ArgTyrLySerProLeuValAlaAlaGluAlaValSerArgGlnGlyTyrAspAlaAlaVal 154
DB 11646 CGGTATGTCACCTGCGTCCGACCAAAATATATACGCAACAGGAATATGATCAGCGATT 11705
QY 155 ThrAlaLysArgSerAlaGluAlaGlyValLysAlaAlaGlnAlaAlaIleLysSerAla 174
DB 11706 GCGGACCGCCGACGAGCGATGCGCGCTGTGCGCGCAAAAGCCGCTGTGAAAGCGCG 11765
QY 175 GlyIleAsnLeuAsnArgSerArgIleThrAlaProIleSerGlyPheIleGlnSer 194
DB 11766 CGATTATTAACCTTCGTATATACCAAGTACCTCACCCATCAGCGCGCTATAGGAAATCT 11825
QY 195 LysValSerGlnGlyThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIleArg 214
DB 11826 AATGTGACTAAGCGCGCGCTGTGACTCAATATGTCAGTCACTGACCTGACCTGCGCAA 11885
QY 215 GlnThrAsnProMetTyrValAsnValThrGlnSerAlaSerGluValMetLysLeuArg 234
DB 11886 CAACCTCGATCCGATTATATGTCAGCTGACCAATCAAGCAACGACTTATGCGACTCAAG 11945
QY 235 ArgGlnIleAlaGlnGlyLysLeuLeuAlaAlaAspGlyValIleAlaValGlyIleLys 254
DB 11946 CAATCGCTGCAACAGGTAACTGCAATMAAGACAGCGCCAGTACAGCTTCAACTGTA 12005
QY 255 PheAspAspGlyThrValTyrProGluLysGlyArgLeuPheAlaAspProValVal 274
DB 12006 ATGAAATATGTCAGCTCAATCCGATTAAGACAGCTGCAATTTCCGACGTAAACGTA 12065
QY 275 AsnGluSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeu 294
DB 12066 GATGAAAGCACCGGCTCTATCAGCTCAGCGCGGTGTCTCTTAACCCGCAACAGCTCTG 12125
QY 295 MetProGlyLeuTyrValArgValIleMetCaspGlnValAlaValAspAsnAlaPheVal 314
DB 12126 CTTCGCGTATGTGTTGTCGCGCCGCGCATGATGAAAGCGCTCCAGCCCAATGCTCTT 12185
QY 315 ValProGlnGlnAlaValThrArgGlyAlaLysAsp-----ThrValMetIleValAsn 332
DB 12186 GTCCCCAGAGGCGCTAATCCCGACCGCGCGCGGCGGACCCCAATGTCGATGCTTAAAC 12245
QY 333 AlaGlnGlyLysMetGluProArgGluValThrValAlaGlnGlnGlnGlyThrAsnTyr 352
DB 12246 GATTAAGACGAGGTGCAAGCCGCAATGTGTCGCGCGGACGCTATTGCGATTAATGCG 12305
QY 353 IleValThrSerIleLeuLysAspGlyAspLysValValValGluGlyIleSerIleAla 372
DB 12306 CTCATCAGCGAAGGTTAAACCGGCGCATTAAGTCACTGTCAGCGGCTTA----- 12356
QY 373 GlyIleThrGlyAlaLysLysValThrProLysGluThrAlaSerSerGluAsnGlnAla 392
DB 12357 -----CAAAA 12362
QY 393 AlaAlaProGlnSerGlyValGlnThrAlaSerGluAlaLysThrAlaSerGluAlaGlu 412
DB 12363 GCGCGAAGCGGCGCTCAGGTAAGCAACGACATGCTCTTCGAGGAAAGGCGGCA 12422
RESULT 11
AE005757/c 11110 bp DNA linear BCT 12-JUN-2002
LOCUS Caulobacter crescentus CB15 section 83 of 359 of the complete
DEFINITION genome.
ACCESSION AE005757 AE005673
VERSION AE005757.1 GI:113422051
KEYWORDS
ORGANISM
SOURCE
Caulobacter crescentus CB15
Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;
Caulobacteraceae; Caulobacter.
REFERENCE
1 (bases 1 to 11110)
Nierman, W.C., Feldblyum, T.V., Laub, M.T., Paulsen, I.T., Nelson, K.E.,
Bisen, J., Heidelberg, J.F., Alley, M.R., Ohta, N., Maddock, J.R.,
Potocka, I., Nelson, W.C., Newton, A., Stephens, C., Phadke, N.D.,

```

Ely, B., DeBoy, R. T., Dodson, R. J., Durkin, A. S., Gwinn, M. L.,
 Haft, D. H., Kolonay, J. F., Smit, J., Craven, M. B., Khoui, H.,
 Shetty, J., Berry, K., Uterback, T., Tran, K., Wolf, A., Vamathevan, J.,
 Ermolaeva, M., White, O., Salzberg, S. L., Venter, J. C., Shapiro, L. and
 Fraser, C. M.
 Complete genome sequence of *Caulobacter crescentus*
 Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)
 11259647
 2 (bases 1 to 11110)
 Niemann, W. C., Feldblyum, T. V., Paulsen, I. T., Nelson, K. E., Eisen, J.,
 Heidelberg, J. F., Alley, M. R. K., Ohta, N., Maddock, J. R., Potocka, I.,
 Nelson, M. C., Newton, A., Stephens, C., Phadke, N. D., Ely, B.,
 Laub, M. T., DeBoy, R. T., Dodson, R. J., Durkin, A. S., Gwinn, M. L.,
 Haft, D. H., Kolonay, J. F., Smit, J., Craven, M., Khoui, H., Shetty, J.,
 Berry, K., Uterback, T., Tran, K., Wolf, A., Vamathevan, J.,
 Ermolaeva, M., White, O., Salzberg, S. L., Shapiro, L., Venter, J. C. and
 Fraser, C. M.
 Direct Submission
 Submitted (31-JAN-2001) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA
 Location/Qualifiers
 1. .11110
 /organism="Caulobacter crescentus CB15"
 /mol_type="Genomic DNA"
 /strain="CB15"
 /db_xref="taxon:190650"
 254. .721
 /gene="CC0805"
 254. .721
 /note="identified by match to protein family HMM"
 /codon_start=1
 /transl_table=11
 /product="leucine-responsive regulatory protein, putative"
 /protein_id="AAK22790.1"
 /db_xref="GI:13422052"
 /translation="MAISALDDTDRLKLVOSDGRITNAELAKCNLSPATDRYR
 RLRSRGFTGYAALLDPKAVDRALLIPEVVLLETTGDPFEFGAAYRRAPELECHM
 VAGSPYLKARVADMEKAYRSLFDILVKMGVETRYVAVLEBKSTVOLPL"
 complement(1149. .2552)
 /gene="CC0806"
 complement(1149. .2552)
 /gene="CC0806"
 /note="similar to GB:123839 PID:404773 PID:833173;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="efflux system protein"
 /protein_id="AAK22791.1"
 /db_xref="GI:13422053"
 /translation="MLRNFTLILVASTVAGCTIAPQYRPVLPVAQTWSTPAEBAAT
 TVASADLWQVLDVRLQGVVDLALQONRLRVAVLIERAKAQYGVORALPGVS
 GAVGEORGRTPASTPGQPEPTDFVFNATIGFTAYELDLFGRVSLRDLAALQSLATE
 QTSSTSVSLIAEIAANAMLTAAQOERLAKTTLATEEDLRLVROBVDAGSLDPA
 LRNETTAAETRADAVAVYTAQVODRNALTLAGDLPVDLLPGRLOADSVLIDPA
 GLPSDVARPPDVLAEROLQANANATGAARPPRLTSGASGASSTDDIGFKSG
 TSISFTFQISLPIFAGGANYANLISGADDIIVATYETKVTQTAFFREVSALSRAT
 VADRLTAQERLTAAADSTRLSQQRDGLDLSALTLLDAQRSYTAQGLISARLAQA
 TNLVLYKTLGGGAPAT"
 complement(12554. .5709)
 /gene="CC0807"
 complement(12554. .5709)
 /gene="CC0807"
 /note="identified by match to protein family HMM"
 /codon_start=1
 /transl_table=11
 /product="AcetA/AcrD/AcrF family protein"
 /protein_id="AAK22792.1"
 /db_xref="GI:13422054"
 /translation="WLSRFTDRPIFAWVIAIVIMLAGAIAIRLPINQVPEIALPOV
 SISANYPASAKTVDSVTVQVIEQMGKLDGLDWSSTSDSGSATVTLTFKAGTDID
 IAAOVQVKLOTATALLPQEVQOQGLTVAKSARFMVMTIGLSEDPKTTGADLDVMA

SNLQDPLSRVDVGDIQLFGSOYAMRIWLDPOKLAESLTPADVAAAIRAQNAVSG
 QIGCTENLPCTGANATITQASRLQTEQFOPIIVKRTSGCATLDDVAVELGAEYS
 ASIAKNGXOAGMAIKLAFGANALTAATVAAVNRMMQENGPANRYKVPYSPFV
 KLSIEVTVITVAITLIPFVIMFLPIQNRATLIPITAVVILGTEGVLAAGSIN
 TLMFGVLAIGLVDAIVYENVERVNSBESLSEKARSKMETKGLICIAVLV
 AAVFPMAPFGGQGVYKOFSTITVSAAVSVVALLVLPALATMLKPKVKGHEE
 KTFPFMFMNRSFMDMSRYQSVRKILGKRGRWMAIYAAIIVMGLFVLPSPAFPE
 EDQGTMTVLQOLPAGATEEKTLLVLDQVRDHPFLVGEKAVQSVFTVSGFSFAGQNA
 GLAFVRLKDEHRSANLKAQAVAGRAMGAFSOFQDMVPAIVPAVOELGSSGDFP
 QLODVGCVGHEALNARNMMLGMAQSDPSLVGVRPNQODPTPOLKIEVDQKAGGL
 TTPDINSALSAAGSGSYNDFTDRCYVKRYMADAPRPMPTDINRMYRNNQGV
 AFAFPAKSWTYSRDERNGISSVNIQCAPRPGSSGDAIAMEKIAKLPGCVY
 EWMGAQELAEANQAPALVAISILVFFLLALYBSWSIPLAVIMVPIUGVALLA
 TFRGGLSNDIYFQVGLITWGLAKKAILIVERAKDLYEKGMGLIEATLEAVRLRP
 IMTSLAFVGVYPLAISNGAGSQAQHAIGTIVIGMISATLLAIFVPLFVVEKI
 FKXKHGHDAQGVGSTERH"
 complement(5722. .6888)
 /gene="CC0808"
 complement(5722. .6888)
 /gene="CC0808"
 /note="identified by match to protein family HMM"
 /codon_start=1
 /transl_table=11
 /product="HlyD family secretion protein"
 /protein_id="AAK22793.1"
 /db_xref="GI:13422055"
 /translation="MHPQREVAATTVGLAIALITLACGCGPGSGMGMGMPTEVY
 IYVQSSVGLSTELAGRTSYLVSEVRPOVSQVYQKRLPEEGAVVAGQPLVQIDPAT
 YQAAVNSAALALQAOQATPAKLDKADRTALVETGANSRQDDNDQALALQTLAAVG
 YQAAALDSARININVARVTAIPISRIKSSVTAPLGAIVTANQALATVQDLSYVLD
 TOTSAELKLOAQOPASGKVSQAVTLLLEGSTYPIGRLEFSDITVDPGGAAG
 LRVAFPDKGVLLPGMYRAVLISQVAPSGIILPQAVNNDPGKATVMVYCAKGP
 RPYLTQGTVDGDKLVLTSGNADGKVIYEGMKVRPGAPIAVPAGAAPAQAR"
 7182. .7670
 /gene="CC0809"
 7182. .7670
 /gene="CC0809"
 /note="identified by match to protein family HMM"
 /codon_start=1
 /transl_table=11
 /product="transcriptional regulator, Tetr family"
 /protein_id="AAK22794.1"
 /db_xref="GI:13422056"
 /translation="MAEIAKAGLSVQOIVRYENKKAIIATAVADPLAEMRDKFAEL
 ERPGTLRDSIDHLPKALIDKCFDERRALALVVAEARNPKVAAVIRADAOERL
 ARAMBERDRKPESEAFHARCEVGLTFGLLRGVNHRDLDRDLTDLVLRSTVRL
 LS"
 complement(7667. .8449)
 /gene="CC0810"
 complement(7667. .8449)
 /gene="CC0810"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAK22795.1"
 /db_xref="GI:13422057"
 /translation="MRTIYTDAPDNTNQLASSRYLAADQDFLSDSRGVRFILE
 FPADEALRWMGVSTTVYFGSARVERDGEPEERARVTAERLRGRIVSEKGAHQNG
 DKVRDNVIATGGGPGLEAANRGAFFVGASIGFNTILPHEOVNPNPCTELTFRFY
 FAKRKMLARANALVFPFGSGTFDELFEILRLRQDKAPPIPIVLVDEAYRSIIN
 FEALVEHGMTAASDLBELVHPADDAEAVMDKLVDCGLPLPGV"
 8564. .10156
 /gene="CC0811"
 8564. .10156
 /gene="CC0811"
 /note="identified by match to protein family HMM"
 /codon_start=1
 /transl_table=11
 /product="methyl-beta-lactamase family protein"
 /protein_id="AAK22796.1"
 /db_xref="GI:13422058"
 /translation="MTISLTFHGAAGCVTGCARLVTPRTTVLIDCGMFGQSKTKAL

NYOPPDAGQVDAVLTHAHIDHSGLLPKMLAGFEGPIYATATDGLAVMLVDAG
 GIGNEVEVSLRNRLRGFPPEVPIYTERDERVAKOPRVKAGEPVIAIDIVATW
 PAGHMEVASIALDVGPEQAKILFSGDLGSGSPFPDDADAGIDHLESTYGD
 RDRGTGDSARKOOLAEVRRAHAAGGGLPAPFAVESOBLLDLDLMGSGVPG
 DIFLDSPLATATVFOERGMWNPOTONPFPALRPSOHLRLDPSADSDGLRKHM
 VITLASGKCDGRKRLQRLIMREAVILCGTOATGTLRLADGRRNTTIGDEI
 SVARIRQLDAYSGLADAGLVAKAKDLPTGVSFVLHAGSPDTEGRRRLVADPP
 AERLVIPLADGFPALDVGGAQPIGEARPLPAPAPGLKLDWNNORVAFQLAQALAA
 APDRARALKALATLEGRP"
 complement (10153 . 11010)
 /gene="CC0812"
 /complement (10153 . 11010)
 /gene="CC0812"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAK22797.1"
 /db_xref="GI:13422059"
 /translation="MLKPLITAAALPLTAHAQBPFRKSKITLVGDSTVAVNSGM
 GGAFCAHTVTSNIACVNLGRGSSRTYREGSWALAKEMASAGFPQTFVLITLGNH
 DQGRPRSTDLATEPFLNLRPFVEBARAAKAPVLPILRRQFKDGOVYDLAPNA
 QATPKLAETNTPVLDHALSRDAVQAMGPVAMRIAEIPPEPVASASGTTGAM
 FYEPVLPGAVALPPRPSPSPSPPCAMSGFKFVDPYTHDGPGLAFSAMVARALADVA
 PALERNLPP"
 BASE COUNT 1713 a 3613 c 4056 g 1728 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 2,31e-36 Length: 11110
 Score: 770.00 Matches: 180
 Percent Similarity: 60.77% Conservative: 57
 Best Local Similarity: 46.15% Mismatches: 131
 Query Match: 38.14% Indels: 22
 DB: 1 Gaps: 9
 US-09-889-756a-2 (1-412) x AE005757 (1-11110)
 QY 16 AAlaAlaValAlaLeuValIleuSerSerCysGly---LysGlyGlyAAspAlaGlnGly 34
 Db 6846 GCGCGCATGGCTCACTCCCTCGCGCTCGCGCGCAAGAGCGCGGCG-----GGC 6799
 QY 35 GlynProAlaGlyArgGluAlaProAlaProValGlyValAlaThrValHisPro 54
 Db 6798 GGCATGGGCGATGGCGATGGCGGCGCGACGAG---GTCGCGCTACATGTCGCGCCACATCG 6742
 QY 55 GlnThrValAlaLeuThrValGluLeuProGlyArgLeuGluSerLeuAspThrAlaAsp 74
 Db 6741 CAGTCGCGGCGCTGTCACGGAACCTGGCGCGCGCACCTCGGCTATCTGCTCCGGA 6682
 QY 75 ValArgAlaGlnValGlyGlyIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIle 94
 Db 6681 GTGCGCGCGCGATGGCGGCGGCTGTCATCGAAGAGGCGGCTGTCGAGAGGCGCGGCTGTC 6622
 QY 95 ArgAlaGlyGlnProLeuThrValGlnIleAspSerSerThrThrThrGluAlaAsnIleGluSer 114
 Db 6621 CGGCGCGCGCGCTGTCATCGAATCGACCGCGCGCACCTATACGCGCGCTATACAGC 6562
 QY 115 AlaArgAlaGlnLeu-----AlaThrAlaGlnAlaThrLeuAlaValAlaAspAlaAsp 132
 Db 6561 GCGCGCGCGCGCTGTCG 6502
 QY 133 LeuAlaArgThrValProLeuValAlaAlaGlnAlaValSerArgGlnGlnIleThrValAspAla 152
 Db 6501 -----GCTACCAAGGCGCTGTCAGACCGCGCGGCTCTCAAGCGAGACATGACGAC 6448
 QY 153 AlaValThrAlaValArgSerAlaGlnAlaGlyValIleValAlaAlaAlaIleLys 172
 Db 6447 GCCAGGCGCGCGCGCTGTCAGACCGCGCGGCTCTCAAGCGAGACATGACGAC 6388
 QY 173 SerAlaGlyIleAsnLeuAsnArgSerArgIleThrAlaProIleSerGlyPheIleGly 192
 Db 6387 AGCGCGCGGATCTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 6328

QY 193 GlnSerLysValSerGlnGlyThrLeuLeuAsnAlaGlyAAspThrThrValIleAlaThr 212
 Db 6327 AAGAGGACCGCTACCGCGCGCGCTGTCACCGCGCGCAAGCGAGACCGCGCTGCGCACG 6268
 QY 213 IleArgGlnThrAsnProMetThrValAlaAsnValThrGlnSerAlaSerGluValMetLys 232
 Db 6267 GTCCAAACCGCTCGAAGGCTCTATGTCACCTGACCGACGACCTCGCGCGCGCGCGCGCG 6208
 QY 233 LeuArgArgGlnIleAlaGlnGlyLysLeuLeuAlaAlaAspGlyValIleAlaValGly 252
 Db 6207 CTCAGGCGCGCGCTGTCG 6151
 QY 253 IleLysPheAspAspGlyThrValThrProGlnIleGlyArgLeuLeuPheAlaAspPro 272
 Db 6150 CTGAAGCTGGAAGAGCGGCTGACCTATCCGATCCCGGCGCGCGCGCGCGCGCGCGCGCG 6091
 QY 273 ValValAlaGlnLeuSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAspGlnAsn 292
 Db 6090 ACCCTGCATCCG 6031
 QY 293 IleLeuMetProGlyLeuArgValArgValLeuMetAspGlnValAlaValAspAla 312
 Db 6030 GTGCTGCTGCG 5971
 QY 313 PheValValProGlnGlnAlaValThrArgGlyValAlaLysAsp-----ThrValMetIle 330
 Db 5970 ATCCTGATCCCG 5911
 QY 331 ValAsnAlaGlnGlyMetGluProArgGluValThrValAlaGlnGlnGlnGlyThr 350
 Db 5910 GTCCG 5854
 QY 351 AsnThrIleValThrSerGlyLeuLysAspGlyValAspLysValValGlnGlyIleSer 370
 Db 5853 AAGTGGCTGTCACAGCG 5797
 QY 371 IleAlaGlyIleThrGlyAlaLysValThrProLysGlnIleThrAlaSerSerGluAsn 390
 Db 5796 -----ATGAGGTGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5758
 QY 391 GlnAlaAlaAlaProGlnSerGlyValGln 400
 Db 5757 GCGCGCGCTGCG 5728
 RESULT 12
 AE012380/c 13945 bp DNA linear BCT 23-MAY-2002
 LOCUS Xanthomonas campestris pv. campestris str. ATCC 33913
 DEFINITION of 460 of the complete genome.
 ACCESSION AE012380 AE008922
 VERSION AE012380.1 GI:21113851
 KEYWORDS
 ORGANISM
 Xanthomonas campestris pv. campestris str. ATCC 33913
 Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 Xanthomonadaceae; Xanthomonas.
 1 (bases 1 to 13945)
 da Silva, A.C.R., Ferrito, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R.,
 Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida
 Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolino, M.C.,
 Camargo, L.E.A., Camarotte, G., Camavari, F., Cardoso, J.,
 Chambergo, F., Ciapina, L.P., Cicarelli, R.M.B., Coutinho, L.L.,
 Chrisino-Santos, J.R., El-Dorri, H., Faria, J.B., Ferreira, A.J.S.,
 Fereira, R.C.C., Ferrito, M.I.T., Formighieri, E.F., Franco, M.C.,
 Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite
 Jr., R.P., Lemos, B.G.M., Lemos, M.V.F., Locelli, E.C., Machado, M.A.,
 Medeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J.,
 Monck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M.,
 Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A.,
 Rossi, A., Sena, J.A.D., Silva, C.J. de Souza, R.F., Spíndola, L.A.F.,
 Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezera, R.I.D., Trindade dos
 Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and

TITLE Kitajima, J.P.
Comparison of the genomes of two Xanthomonas pathogens with differing host specificities

JOURNAL Nature 417 (6887), 459-463 (2002)

MEDLINE 22022145

PUBMED 12024217

REFERENCE 2 (bases 1 to 13945)
da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R., Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida Jr., N.F., Alves, L.M.C. do Amaral, A.M., Bertolini, M.C., Camargo, L.E.A., Camarotte, G., Camarvan, F., Cardozo, J., Chusinberg, F., Clapina, L.P., Clacarelli, R.M.B., Coutinho, L.L., Cusinho-Santos, J.R., El-Dorri, H., Faria, J.B., Ferreira, A.J.S., Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C., Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locati, E.C., Machado, M.A., Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Melandri, J., Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M., Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A., Rossi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F., Taita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and Kitajima, J.P.

FEATURES Direct Submission
Submitted (28-NOV-2001) Departamento de Bioquímica, Universidade de Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900, Brazil

TITLE JOURNAL

SOURCE 1. 13945
/organism="Xanthomonas campestris pv. campestris str. ATCC 33913"
/mol_type="genomic DNA"
/strain="ATCC:33913"
/db_xref="ATCC:33913"
/db_xref="taxon:190485"
/note="pathovar: campestris"
151. 1854
/gene="XCC2674"
151. 1854
/gene="XCC2674"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
/transl_table=1
/codon_start=1
/product="extracellular serine protease"
/protein_id="AAM41946.1"
/db_xref="GI:21113852"
/translation="MLAQTPATSRAPQPPRESDDLFRYQHLNDQEVIAIDTPRA GTDINVGPLHTLGRGGVAVVVDALDELHPDLVANVVGSGKNFQDGSNDTPRN GASHGTWVGIMAAVGNGLGRGVPARLKNLISSESPDPTDNLRAQSDVY OSRDVDNNSFGSDLTYPTISPAEBSILRLMRNRANGKGLYVOAGNTPDSFTY LDDGQNVRECPVARTILGVTCTPATDPLISNOLITATGAVNARGLRSSVSGAL WYTFGGEFGIQRKRYFGDRPALFDPALIVTTDLTGCAVGNRIAGPNNALASS SPIDASCNYSAAWNGTSAAPTVAGVALIVQANPSLARDKYLATISADIPAOQ VIRYQGVIEPGWITNAGHAFSNWYGLVDAEAVYRAYFQPLPORDLQKAR AGTSTIGPNNAATRLRLGDLTKIDTVQSMQTHKPSNLRVLLISPSGTSRYLT PQALDITLQAGFEIPLSTSNAPLDENVAGTWLLEVTDMGSDVPAGLTFKLRING H"

gene 1856. 2446
/gene="XCC2675"
1856. 2446
/gene="XCC2675"
/note="putative; ORF located using Glimmer/Genemark"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAM41947.1"
/db_xref="GI:21113853"
/translation="MNIALGTLILCALASMPASAOPTDRLAORAAQGGFTIGDN FLVFNARVLENTTPPRADVPANRSTTAAGTTLGHICGYVSSGNSPPTTYP GIVEPSLMAVINGNPGPVASPLQVYADADAIVKLAQTNGAVLAASQDGTGYI GYRSVPEALALPTVRGGVDVLTQYQVPLRLR"

CDS 3031. 3942
/product="transcriptional regulator"

gene 3031. 3942
/gene="XCC2676"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
/transl_table=1
/codon_start=1
/product="transcriptional regulator"
/protein_id="AAM41948.1"
/db_xref="GI:21113854"
/translation="MASSIMHLIDVQRNAPDTGTPSTAVPGRLMRCAPTSYIPV YTPCLVAGSHQOMLGTAVHHPAMTVASVDLITGSVIEASPAQYCLCCLD DVAVLSLALRYPALAEQADTPASGLILNRSTPELMDALRLARDPRDIDALP LVIRELLRYMAEFGNAIVEQMAIADSRILQISPAIMLVAERFSIQIDLSAM SRSTFHHFAVMSMTPLERFSRLRVEARLIMVAELSAADAGFRVYESASQFSRD YARICGPPARDQQLRASRODAIEVA"
complement (4154. 4591)
/gene="XCC2677"
complement (4154. 4591)
/gene="XCC2677"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAM41949.1"
/db_xref="GI:21113855"
/translation="MAAGRGADENHAPIKYTKAGSASAGAPSYFTGTVRIDAPPOS DAPRVGATVTFEPGARITKWHHTPLQOTLIVTAGRGVDEKGRPAOEIRPGDVLWLP PGVHMHGANAIVAMSHIALAEAVDSPTWLEQVSDAQGN"
4794. 5708
/gene="XCC2678"
4794. 5708
/gene="XCC2678"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
/transl_table=1
/codon_start=1
/product="transcriptional regulator lysr family"
/protein_id="AAM41950.1"
/db_xref="GI:21113856"
/translation="MWSASWARENLDLQVFTVAREGSTRAAAQGVQSALSHTV RALETRIGIRLITRTSRVSMTEAGMKLPFTVAPRLEIQEBSALTPAFDKPAGTIR ITAGHAAADVILWPAISKVLPDYDEITVDVGLADIVAEKYDIGRLGDQVAKDM IAVDISPROMAVVAASRYFVHTTIPAVPRLAONHNGIRLPTFGHLLPMDFFKQDR EOKRVNDGWTENGSAMLRBALAGLAPLPTWLDVHIAAGLRVLDMDKCPDFG YIAYPSRQSSAAKVTIDALRHDLVR"
complement (6036. 7019)
/gene="XCC2679"
complement (6036. 7019)
/gene="XCC2679"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=1
/product="oxidoreductase"
/protein_id="AAM41951.1"
/db_xref="GI:21113857"
/translation="WHTRTLGRSGLRVSALGCMGLSYGYPATERKQAVALLHAAY EOGVTLPTDAEAYGPINLELGEBALPYGAQVVAATKFGROHADAGLDSRAPAR AVAEASILRLKRERIDLFYORVDPAVPIEIVAGTVADLTAEGVGHFGJSEASATIR RRAHAYQVAAVQSEYGLIMREPEASVLPLEEIGIGIFVPSPLGKALFGALQOPDQ FQDSDPNSVPRFAAEARBANQALVQGLIADDKGTPPAQVLAWLMLSKRPVLPVLP GTTKRRLDENIGGALVSSDELARIQALDAVEIVGERXSPERQLVK"

gene complement (7291. 7980)
/gene="XCC2680"
complement (7291. 7980)
/gene="XCC2680"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=1
/product="transcriptional regulator"

9606, 27 Bd Jean Moulin, 13385 Marseille cedex 05, FRANCE

FEATURES
source

1..6066
/location="Qualifiers"
/organism="Enterobacter aerogenes"
/mol_type="genomic DNA"
/strain="BML627"
/db_xref="taxon:548"
complement(1..506)
/gene="aeaf"
complement(<1..506)
/gene="aeaf"
/function="putative transmembrane protein"
/codon_start=1
/transl_table=1
/product="putative AefA protein"
/protein_id="CAC35722.1"
/db_xref="GI:13539228"
/db_xref="SPTREMBL:O9AEG4"
/translation="MOHTRWPHASTIVFALVIVGTLSPFQVAAADLPDRSDIOG
LTTANKEKELPQEKLVQOQDLTQTLTDLKIERIKSEVQLRQOEAPAMQAVDS
LNAISVDVDEVTREKTLSTLSLRQLESRSVQALDLOSAQNDLATVNSQLVLTQPE
RVQNAWSA"
complement(637..1287)
/gene="acrR"
complement(637..1287)
/gene="acrR"
/function="putative repressor of acraB operon"
/note="Hcr family regulator"
/codon_start=1
/transl_table=1
/product="putative AcrR protein"
/protein_id="CAC35723.1"
/db_xref="GI:13539230"
/db_xref="SPTREMBL:O9AEG3"
/translation="MARKTKQALFTQQLIDVALRLFSQGSSTSLAATAKAGVT
RGATVYHFKNSDLFNEIWSLSDASISDELYEAKFNDPLSVRELIVTLEIATV
EERRLMWELIHFKECEFGEMAVVQOQSLMSYRIEQLDQCTTAQDLNLLT
RRAAILMRSYSGMENWLFAPESFNHAEARAVVDLIEWYQCPSPHSSSEVMA"
1419..2629
/gene="acrA"
1419..1422
/gene="acrA"
1430..2629
/gene="acrA"
/function="component of an efflux pump"
/codon_start=1
/evidence="experimental"
/transl_table=1
/product="acrA protein"
/protein_id="CAC35724.1"
/db_xref="GI:13539231"
/db_xref="SPTREMBL:O9AEG2"
/translation="MNRKRGITPLAVVIMLSGSLATGCDKPKAQQCAQCPVCIYT
LKSAPLOTTELPGRTNAVRYAEVRPQVSGIILKRNFTGEGSDIAGVSLQIDATVQ
ASVSAGKDLAKQAANAIAQITVRYOKLGTXYISQOEYDVAADQOSNAVVA
KAAYETARINLAKYTPISIGRIKSAVTEGALVQNGSTALATVQDLDPIDVDT
SSNPELRKQELANGKLENGKAKVELVTNGKYRQEGTLESDVTVDDTGSITL
RAIPNPDPHTLLPGMFVPAKLEBGINPDALLIPQOQYTRTPRGASVWVVEGPKVY
KQVTAQAIQKMLVTBGLKSGDRYVAGLQKAPGVQVAKQAEVASDHOQTAGNANA
QSEQTKS"
1430..1501
/gene="acrA"
1502..2626
/gene="acrA"
/gene="acrA"
/product="acrA"
2637..5833
/gene="acrA"
/gene="acrB"
2637..2641
/gene="acrB"
2652..5798
/gene="acrB"
/function="inner membrane component of an efflux pump"
/codon_start=1

/evidence="experimental"
/transl_table=1
/product="acrB protein"
/protein_id="CAC35725.1"
/db_xref="GI:13539232"
/db_xref="SPTREMBL:O9AEG1"
/translation="MNFIDRPIFAWVIAIIMLAGSLIMKLPAQVPTIAPPAVT
ISATVPGADAKTVQDTVQVIBQNMNGIDIMLWMSSDSTGVQVLTLPQSGDADI
AQVOQNKLOLAMPLLPOEQOQSVSESSFLWAVGVINTGWTQEDISDYVA
MKDASRTSGVDVOLFGSOYAMRIIMDDTKNNPOLTATVDVINALKAQNAOVAAG
LGSTPYKGOOLNASTIAQRLTSADSEFSIILKKNQSOVARLPDAKVELGEGND
IANKNQPSGIGIKILATCANLIDTANALRALAKNEPIFSGKLIVYDTPPEK
ISLHEVVKTLVEAILIVFLWVYLFQNFATLLPTIAVPAVLLEFPAIIVFGSLIT
LTFMGVLAIGLIVDAIIVVENVERVMEEGLPKPKATRSKQOIGALVGIAMVLS
AVFIPIPAFPGSGTAIVROFSTITVAMALSVLVALITFALCATMLKPIQKSGHSH
KGFPGFNMRKDSSTHYDTSVGNILRSTGRVYLIIIVGNAVSFVPAVNGFGRSGNTP
DOGVFLSMQLPAGASQOEPRQVLDQMTQVYLTKKNRNVSVPAVNGFGRSGNTP
IATVSLKMSERPGSEKKEAITIGRAMARSQIKDANVFENLPAIYELCTATGFPDQ
LIDQGGHGHKLTOARNOQLFGVIAOHFDLLVGRPNGLDETPOPKIDIDOKQALGV
SIDINTTLTAAGGSYVNDPIDRGVKKVMSSEAKRYMLPEDIGMAYRSGDQNV
PFSAFSTSHMEYSPRLERYNGLPSEMIIGQAAFGSTGBAMMEQLASKLPSGVLY
DMTGMYSOERLSGNOAPALYALSLIVFLALAYESWSIPFSLVLPVPIGVALIA
ATPRGLNDVYPOVGLITTGLSAKNAIILIVEPAKDMREKGLTEATEAVPMR
PIIMTSLAFLIGVMPPLVITSGAGSGAQNAGVTGMSGATATVLAIFVFVFPVVR
RFTKKTEDIHNPVEBH"
5812..5833
/gene="acrB"
terminator
BASE COUNT 1442 a 1574 c 1628 g 1422 t
ORIGIN
Alignment Scores:
Pred. No.: 2,56e-36 Length: 6066
Score: 763.50 Matches: 169
Percent Similarity: 61.94% Conservative: 80
Best Local Similarity: 42.04% Mismatches: 134
Query Match: 37.82% Indels: 19
Gaps: 5
US-09-889-756a-2 (1-412) x EAE306389 (1-6066)
QY 11 AAlaAlaLeuAlaAlaAlaValAlaLeuValLeuSerSerGlyGlyValGlyAsp 30
Db 1460 GCGGTCGTCGTGATGCTCTCAGGACGCTTACCGCTAACAGATGT--GACGATTAACCG 1516
QY 31 AAlaLagInGlyGInProAlaGlyArgGluAlaProAlaProValGlyValVal 50
Db 1517 GCTTAACAAGAGCCGCG-----CAATGCCGAGATCGGATTC 1558
QY 51 ThrValHisProGInThrValAlaLeuThrValGluLeuProGlyArgLeuGluSerLeu 70
Db 1559 ACGGTCAAAATCCGACCTCTCAAAATCACCACCTGAACTTCCTGCTTACCAATGCTTAT 1618
QY 71 ArgThrAlaSerValAlaGluAlaGluValGlyGlyLeuGlnLysArgLeuPheGlnGlu 90
Db 1619 CGTGTCCCAAGAGCCGCTCTCAGAGTAAAGCTTATTCGAAACGTAACCTTACTGAA 1678
QY 91 GlySerTyrValArgAlaGlyGInProLeuTyrGlnIleAspSerSerThrTyrGluAla 110
Db 1679 GCGACCATATTCAGCGGGGGGCTCTCTTATTAAGTTCATCCGCAACTATCAGGCA 1738
QY 111 AsnLeuGluSerAlaArgAlaGluAlaGluThrAlaGluAlaThrLeuAlaValAsp 130
Db 1739 AGCTATGAAAGCGCTTAAGCGACCTTAGCGAAAGCCAGGCCCGCAACATGCTCAG 1798
QY 131 AlaAspLeuAlaArgTyrLysProLeuValAlaAlaGluAlaValSerArgGlnGluTyr 150
Db 1799 TTGACGGTCAAACTTATCAGAACTGTTGGCACCMAAATACATCAGCAACAAGATAT 1858
QY 151 AspAlaAlaValThrAlaLysArgSerAlaGluAlaGlyValLysAlaAlaGluAla 170
Db 1859 GACTCAGCGTGTCCGACCGCTCAGCAAAAGCACGCCCTCGTGGCGGCAAAAGCCGCA 1918
QY 171 IleLysSerAlaGlyIleAsnLeuAsnArgSerArgIleThrAlaProIleSerGlyPhe 190

REFERENCE 1
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glaesner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K., Apodaca,U., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.
Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
Nature 409 (6819), 529-533 (2001)
21074935
PUBMED 11206551
REFERENCE 2
(bases 1 to 12524)
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glaesner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K., Apodaca,U., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.
Direct Submission
Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
Location/Qualifiers
1. 12524
/organism="Escherichia coli O157:H7 EDL933"
/mol_type="genomic DNA"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/note="enterohemorrhagic"
complement (110. .619)
/gene="y1aC"
/note="synonym: Z0570"
complement (110. .619)
/gene="y1aC"
/function="orf; Unknown function"
/note="Residues 1 to 169 of 169 are 100.00 pct identical to residues 1 to 169 of 169 from Escherichia coli K-12 Strain MG1655; B0458"
/codon_start=1
/transl_table=1
/product="orf, hypothetical protein"
/protein_id="AAG54807.1"
/db_xref="GI:12513325"
/translation="MPEPKIQRPRQMTETIQRLTITSLNTRKDKKPPRSISFIRKPGLPFGIMYVAFATLAVMLQSETLSGWLVLVLLFLLGFFPDVYPRYREDIDVLDPRVCYNGEWYNTRFVPAALVEALINSPRVADVHKEQLKMIVRKGLSFYDIFTLRAESTS"
complement (696. .1247)
/gene="y1aD"
/note="synonym: Z0571"
complement (696. .1247)
/gene="y1aD"
/function="putative enzyme; Not classified"
/note="Residues 1 to 183 of 183 are 99.45 pct identical to residues 1 to 183 of 183 from Escherichia coli K-12 Strain MG1655; B0459"
/codon_start=1
/transl_table=1
/product="putative transferase"
/protein_id="AAG54808.1"
/db_xref="GI:12513326"
/translation="MSTREKMTAGELYSRSDETLSRDRIARQLIHRVHSLAEHTLRQILDLFGQVTEATIEPTFRCDYGINIFLGANFPANDCVMLDVCPTRIDNCMLALGVHLYTATHPIDPVARNSGALGKRVITIGNVMVIGRAVINPGVTIGNVVAVASGAVVTKDVPDNNVVGNPARIITKL"
complement (1419. .1637)
/gene="y1aH"
/note="synonym: Z0573"
complement (1419. .1637)
/gene="y1aH"
/function="regulator; Macromolecule synthesis,

modification: Proteins - translation and modification"
/note="Residues 1 to 72 of 72 are 100.00 pct identical to residues 1 to 72 of 72 from Escherichia coli K-12 Strain MG1655; B0460"
/codon_start=1
/transl_table=1
/product="haemolysin expression modulating protein"
/protein_id="AAG54809.1"
/db_xref="GI:12513327"
/translation="MSRKPLTKTDYLMRLRCQTIDTLERVIEKNKVELSDNELAVFYSAADHRLAELTNMKYDKIPSSVWKFR"
complement (1663. .2037)
/gene="y1aJ"
/note="synonym: Z0574"
complement (1663. .2037)
/gene="y1aJ"
/function="orf; Unknown function"
/note="Residues 1 to 124 of 124 are 100.00 pct identical to residues 1 to 124 of 124 from Escherichia coli K-12 Strain MG1655; B0461"
/codon_start=1
/transl_table=1
/product="orf, hypothetical protein"
/protein_id="AAG54810.1"
/db_xref="GI:12513328"
/translation="MDEYSPKRDIAQLKFCETLYHDCILANESNHNQVNDPTSAINQINLEIHLATFALNLYKIKVNEDKLLIQIDBYLDPTFMLSISGINNDDQKMKSGNRLPRCPVNAKKNPAGLSC"
complement (2583. .5732)
/gene="acxB"
/note="synonym: Z0576"
complement (2583. .5732)
/gene="acxB"
/function="transport; Protection responses: Drug/analog sensitivity"
/note="Residues 1 to 1049 of 1049 are 99.80 pct identical to residues 1 to 1049 of 1049 from Escherichia coli K-12 Strain MG1655; B0462"
/codon_start=1
/transl_table=1
/product="acridine efflux pump"
/protein_id="AAG54811.1"
/db_xref="GI:12513329"
/translation="MNPFRIDRPIFAMVIAIITMLAGGLAILKLPVAYPTIPPAVTISAVPGDAKTYODPTOVITBONNKNINLMWMSNSSTVOITLTPESTDDIAQOVONKIQAMPFLIPQVQOQVSTIEKSSSFLWVGVINTDGMTDSDYVAA NMDAISRTSGVDVOLFGSOYAMRIWMNPNELNKFQLPVDTITAIKQNAQVAAQ LGGTPPVKGOQLNASTIIAOTRLTSTEEFGILKVNQDGSRVLLRVAKIELEGEND IIAEFGPQASGIGIKLATGANALDTPAAIRALAYMEPPFSGIKIIVYDTPPVK ISIHVYKTLVEAIIIVFLVMYFLQNPATLITPTIAVPVLLGTFVLAAGFASINT LTPRGVLAIGLIDVAIVVERVERVMAEBGLPREAKRKSGQIQGALVGIAMTLS AVFVPVAFPGSGTATYROPSTITVSAMLSVLAIVLITPALTATMLKPLAKDHDGSG KGPFQGFNRMPEKSTHYATDSVGLIRSTRGLVYLIIIVMAALFAPLPSFLPD EDQGVFMVVOLEPAGATQERTQKLVNEVHYITLTKKNVESVFAVANGFARCONT GIAFVLIKMDADPGSENVKVEATIMRATFAFSQIKDAMFAFVLPVIVELGTADGFE ELIDQGLHEKKTQARNOLLAEAAGHPMLTSVRNGLEDPPQFRIIDDOEKAQALG VSINDINTLTGAAMGSSYNDPRIDRVRKYVYMSAKYRMLPDDIDGMYVRADQGV VPSAFSSSRMEYSPRLERINQLPSEVITLGOAPKSGTGEALHLEQLASKPLTGG YMTQMSYQERLSGNQPSLYAISLVILCLALVESISIPSVMLVPLVGLIAGL AATFRLITNDVYFOVGLITLTIGLSANALITVEFADLMDKEKGILTEATLDVVRML RPLMTSLAFLIGVMPLVISTGSGAQAQAVGTVGAVTATVLAIFVVPVFVVR RPRSRKNEDIENHHTVDHH"
complement (5755. .6948)
/gene="acraA"
/note="synonym: Z0578"
complement (5755. .6948)
/gene="acraA"
/function="transport; Protection responses: Drug/analog sensitivity"
/note="Residues 1 to 397 of 397 are 100.00 pct identical to residues 1 to 397 of 397 from Escherichia coli K-12 Strain MG1655; B0463"
/codon_start=1

ORGANISM Escherichia coli K12
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 13446)
AUTHORS Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Coliado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J.J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y.
The complete genome sequence of Escherichia coli K-12
Science 277 (5331), 1453-1474 (1997)

TITLE JOURNAL
MEDLINE 97426617
PubMed 9278503
2 (bases 1 to 13446)
REFERENCE Blattner, F.R.
AUTHORS Direct Submission
JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecol1@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
3 (bases 1 to 13446)
REFERENCE Blattner, F.R.
AUTHORS Direct Submission
JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecol1@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
4 (bases 1 to 13446)
REFERENCE Plunkett, G. III.
AUTHORS Direct Submission
JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHER). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using Genemark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (<http://cgsc.biology.yale.edu>). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (<http://www.genetics.wisc.edu>). *** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'p' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

FEATURES

source Location/Qualifiers
1. 13446
/organism="Escherichia coli K12"
/mol_type="genomic DNA"
/strain="K12"
/sub_strain="MG1655"
/db_xref="taxon:83333"
1. 22
/note="factor Sigma70; predicted +1 start at 475839"
86. 439
/gene="ybaA"
/note="synonym: b0456"
86. 439
/gene="ybaA"
/function="orf; Unknown"

gene
promoter
CDS

/note="orf; 100 pct identical to YBA_ECOLI SW: P09161"
/codon_start=1
/transl_table=1
/product="orf, hypothetical protein"
/protein_id="AAC73558.1"
/db_xref="GI:1786661"
/translation="MKYVDFVVAVPADKKDAYREMAKAPLFEKFGALRIYECMAS
DVPDGKVDPRMAVKAENEVVFWSWIEYPSKEVRDAAQKMSDPMKEGESMPD
GKMIVCGFESIIDE"
complement (481. 2037)
/gene="y1aB"
/note="synonym: b0457"
complement (481. 2037)
/gene="y1aB"
/function="orf; Unknown"
/note="f518; This 518 aa ORF is 33 pct identical (6 gaps)
to 513 residues of an approx. 536 aa protein y0CC_ECOLI
SW: P32701"
/codon_start=1
/transl_table=1
/product="orf, hypothetical protein"
/protein_id="AAC73559.1"
/db_xref="GI:1786662"
/translation="MLVRIHLVGLISGLIIVLPLPGLSLIAHQVETSPIELD
TVSSRAIRANKVATQKALQELERQACSAHMEMRVSYSRYIQEVAYIDN
NVPQCSLIEHSPDTPPEBPQKISDQSVRLMGIIRVWAGTAYVIMIP
ASFDIVPYSSMOIDAALIGNAVVTSSDEIAQGIITLQKTGSHENGVIYDI
LPLPENNISITWASTKMLQKGMHVRVILPGLIVGLAAMPVRLIRIOSPHR
LQALINRDIQVHQPVSILANGKIYGAELARPPODGSMLSDSTIPLAQQGLSE
PLTLILRSVFEDMGMLRQHPQOHISINSESVLTSEKIPLQLRDMINHYOVNROI
ALBLTERFADPRTSAPILISRYRAGHEIYLDPGTQSLSYLQDIDVILKIDKSF
VDALRYENKVPPIHIAKTLKLVKVAEGIETSKQEWLROHGVHYGCGMLSKALPRE
DFRMAEOHL"
complement (2052. 2080)
/note="factor Sigma70; predicted +1 start at 477855"
complement (2195. 2704)
/gene="y1aC"
/note="synonym: b0458"
complement (2195. 2704)
/gene="y1aC"
/function="orf; Unknown"
/note="f169; This 169 aa ORF is 32 pct identical (0 gaps)
to 53 residues of an approx. 592 aa protein ASN1_PBA SW:
P19251"
/codon_start=1
/transl_table=1
/product="orf, hypothetical protein"
/protein_id="AAC73560.1"
/db_xref="GI:1786663"
/translation="MPPGKLQIPRROTMTIORLLTETTESLNTREKRNKPPRSISF
IRHPLPLIGMVAPFATLAVMIQSTLSGSWLVLVPLTILNGPFPDVPYPRRYED
IDVLDPRVCNGBWNTRVFPALVAAILNSPRAVDVHKEQLKMLVRKELSPYDIF
TLARASTS"
complement (2708. 2735)
/note="factor Sigma70; predicted +1 start at 478511"
complement (2722. 2750)
/note="factor Sigma70; predicted +1 start at 478525"
complement (2781. 3332)
/gene="y1aD"
/note="synonym: b0459"
complement (2781. 3332)
/gene="y1aD"
/function="putative enzyme; Not classified"
/note="f183; This 183 aa ORF is 65 pct identical (2 gaps)
to 183 residues of an approx. 192 aa protein YVAI_BACSU
SW: P37515"
/codon_start=1
/transl_table=1
/product="putative transferase"
/protein_id="AAC73561.1"
/db_xref="GI:1786664"
/translation="MSTKEKMIAGELYSADETLNRDLRARQLIHRYNHSLAEHT
LRQQLIADFGQVTEAYIEPTFCDDYNIPLGNPFANFDCVMDVCPRIIGDCML

promoter
gene
CDS

ORGANISM *Escherichia coli*
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 139818)
 Robert, D., Allen, E., Araujo, R., Aparicio, A., Chung, E., Davis, K.,
 Duncan, M., Federspiel, N., Hyman, R., Kaiman, S., Komp, C., Kuri, O.,
 Lew, H., Lin, D., Namath, A., Oetner, P., Schramm, S. and Davis, R.W.
 Sequence of minutes 4-25 of *Escherichia coli*
 Unpublished

TITLE 2 (bases 1 to 139818)
 Robert, D., Allen, E., Araujo, R., Aparicio, A., Chung, E., Davis, K.,
 Duncan, M., Federspiel, N., Hyman, R., Kaiman, S., Komp, C., Kuri, O.,
 Lew, H., Lin, D., Namath, A., Oetner, P., Schramm, S. and Davis, R.W.
 Direct Submission
 Submitted (19-DEC-1996) Department of Biochemistry, Stanford
 University, Stanford, CA 94304, USA

FEATURES

source location/Qualifiers
 1..139818
 /organism="Escherichia coli"
 /mol_type="genomic DNA"
 /db_xref="taxon:562"
 /map="minutes 9 to 11"
 1..684
 /note="overlaps the sequence in GenBank Accession Number
 U73857, which covers the 6 to 8 minute region"
 <1..360
 /note="similar to *S. typhimurium* branched chain amino acid
 transport system II carrier"
 /codon_start=1
 /transl_table=1
 /protein_id="AAB40157.1"
 /db_xref="GI:1773085"
 /translation="LGGFMSVNSLGLSOLISVPLTATYPCIALVVLSPFRSW
 HNSRVLAPMFTSLRGILDGIRASAFSDILPBMQRLPLAEGLAMLPYVMVL
 AITVDRAGROVSSAH"
 436..1809
 /note="similar to *S. typhimurium* Prox"
 /codon_start=1
 /transl_table=1
 /protein_id="AAB40158.1"
 /db_xref="GI:1773086"
 /translation="WESKMKLRKGLSTRHIFMAAGSAGITGLFYGSADAIKAGPSV
 LADVITGIAVYIIMRALGEMSVNPPASSFSRYAENLPLAGYITGYTCFEILIV
 AIAVITFGIYGVWPTVPHMIVLSVLIICAVNLSYKVGELTFMFSFKVAT
 IIMVAGEGIIIMGIGGQPTGTHNLSNGSPFNSGLMMSLQMWMAVYGIETI
 GITAGEKDEKSIIPRAINSVPWRIYFVYGLFVMSIYPMNOVGTAGSPYLITROH
 MGTFRASITINPVUTASLSAINSDVGVCRMLHMAEQSAPKIRSKTSRGTIPWT
 LVTWTTALFAVYLYNTPENVEPLVSLATPATVWVWIMILSQAIFRRRLPEEVK
 ALKRVPGVATTIGLIFLFTIGLIGYHPDTRISLYGVFAWIVLLIGMFKRRHD
 ROLAENG"
 1680..4024
 /note="previously sequenced by another group and can be
 found in GenBank Accession Number X59839"
 1965..3782
 /gene="malZ"
 1965..3782
 /gene="malZ"
 /codon_start=1
 /transl_table=1
 /product="maltodextrin glucosidase"
 /protein_id="AAB40159.1"
 /db_xref="GI:1773087"
 /translation="MMNLAMHLPPRPVKOSKOLLITLMTGDEPPORTIMLTEDHN
 EEMSVPMHQRSGQPGVTAWRAAIDSSQPRRRYSFKLMDHROKMFIPQPSKMP
 PALEBPADVDPDIGQWADQIFYOIFPDRFARSLPREAQDHVYHHAQGEILIR
 DWDEPVTAQAGSTFYFGDLDGISEKLPYIKKGVLYLNPVKAPASVAKYDTEYR
 HVDPQFGDGLRLRHTQOLQMLVLVDGVFNHSGDSHAMFDRHNGTGAQHNES
 PMRWDSFSDGTALDWLGYASLPKLDYQESLVEINIEGSDSIVRHMLCAPNMGM
 RLIVVHMLGEAGGRNNMHVAGITTEAKETOPEAYLVGHPGARQWLOADVEDAM
 NYGFTPELPGPLANDISYDPOIATQCTMAMNDYRGLSHQOOLRMNQDSDHT
 ARKTIILGRDIARPLAVVWLTTPGVPCITYGEGVLGDSKDPCKRPPWVEKOD
 TALFALYRMTALRKKSQALRHGSCVLTVAEDNVVVFVRLVNOQRLVALINRGEACEV
 VLPASPLPNAVOWCKRGHQLTDGLALPAISATVWMN"

misc_feature
 /note="previously sequenced by another group and can be
 found in GenBank Accession Number X59839"
 1965..3782
 /gene="malZ"
 1965..3782
 /gene="malZ"
 /codon_start=1
 /transl_table=1
 /product="maltodextrin glucosidase"
 /protein_id="AAB40159.1"
 /db_xref="GI:1773087"
 /translation="MMNLAMHLPPRPVKOSKOLLITLMTGDEPPORTIMLTEDHN
 EEMSVPMHQRSGQPGVTAWRAAIDSSQPRRRYSFKLMDHROKMFIPQPSKMP
 PALEBPADVDPDIGQWADQIFYOIFPDRFARSLPREAQDHVYHHAQGEILIR
 DWDEPVTAQAGSTFYFGDLDGISEKLPYIKKGVLYLNPVKAPASVAKYDTEYR
 HVDPQFGDGLRLRHTQOLQMLVLVDGVFNHSGDSHAMFDRHNGTGAQHNES
 PMRWDSFSDGTALDWLGYASLPKLDYQESLVEINIEGSDSIVRHMLCAPNMGM
 RLIVVHMLGEAGGRNNMHVAGITTEAKETOPEAYLVGHPGARQWLOADVEDAM
 NYGFTPELPGPLANDISYDPOIATQCTMAMNDYRGLSHQOOLRMNQDSDHT
 ARKTIILGRDIARPLAVVWLTTPGVPCITYGEGVLGDSKDPCKRPPWVEKOD
 TALFALYRMTALRKKSQALRHGSCVLTVAEDNVVVFVRLVNOQRLVALINRGEACEV
 VLPASPLPNAVOWCKRGHQLTDGLALPAISATVWMN"

gene
 /note="previously sequenced by another group and can be
 found in GenBank Accession Number X59839"
 1965..3782
 /gene="malZ"
 1965..3782
 /gene="malZ"
 /codon_start=1
 /transl_table=1
 /product="maltodextrin glucosidase"
 /protein_id="AAB40159.1"
 /db_xref="GI:1773087"
 /translation="MMNLAMHLPPRPVKOSKOLLITLMTGDEPPORTIMLTEDHN
 EEMSVPMHQRSGQPGVTAWRAAIDSSQPRRRYSFKLMDHROKMFIPQPSKMP
 PALEBPADVDPDIGQWADQIFYOIFPDRFARSLPREAQDHVYHHAQGEILIR
 DWDEPVTAQAGSTFYFGDLDGISEKLPYIKKGVLYLNPVKAPASVAKYDTEYR
 HVDPQFGDGLRLRHTQOLQMLVLVDGVFNHSGDSHAMFDRHNGTGAQHNES
 PMRWDSFSDGTALDWLGYASLPKLDYQESLVEINIEGSDSIVRHMLCAPNMGM
 RLIVVHMLGEAGGRNNMHVAGITTEAKETOPEAYLVGHPGARQWLOADVEDAM
 NYGFTPELPGPLANDISYDPOIATQCTMAMNDYRGLSHQOOLRMNQDSDHT
 ARKTIILGRDIARPLAVVWLTTPGVPCITYGEGVLGDSKDPCKRPPWVEKOD
 TALFALYRMTALRKKSQALRHGSCVLTVAEDNVVVFVRLVNOQRLVALINRGEACEV
 VLPASPLPNAVOWCKRGHQLTDGLALPAISATVWMN"

CDS

misc_feature 2538..5551
 /note="previously sequenced by another group and can be
 found in GenBank Accession Number M37702"
 complement (3787..4368)
 /gene="yajB"
 complement (3787..4368)
 /gene="yajB"
 /codon_start=1
 /transl_table=1
 /product="hypothetical 14.0 kd protein"
 /protein_id="AAB40160.1"
 /db_xref="GI:1773088"
 /translation="MNFLEMLHLAHAESLSGNLLADFYKGNPESFPDDVAGIHM
 HRRIYDLTDNLPREAREWFRSETRVAPITLDVMMDHPLSRMSQSLSPDLOEFV
 CVREOVMTLPPSPRPRLINNYLMSEQLVRYRMDFIQVNLNGASRRPLDLRL
 DSWYDDAHAYDALETRFWOYPRMMAQASKAL"
 4461..5531
 /gene="queA"
 4461..5531
 /gene="queA"
 /codon_start=1
 /transl_table=1
 /product="S-adenosylmethionine:tRNA
 ribosyltransferase-1 isomerase"
 /protein_id="AAB40161.1"
 /db_xref="GI:1773089"
 /translation="MRVTDPSFELPSLIIAHYMPERRSSGRLSLDGPGLTHGPT
 DILDKNPGLIYFNNTRYPARLFGKRSAGKIEVLYERMDDKRLIAIRSKAPK
 PGEILLGDDESINATMTAHGALFEVEREDERSVLDLINSIGHMPLPYITDDEDA
 DRELQYQVYSEKPAVAAPTAGHPEPLEKIRAGVENAFTLVHGAQTFQVPRVD
 TIEDHIMSEYAEVPOVDVAVLAARGRVIAQVTSVRSLSAQAAGNLIIEPF
 FDTDFEIFYEQYKVDALVTNPHLPESTLIMLVSAFAGYQTNNAKYAAVEKXRF
 FSGDAMFLITYNQALNERVE"
 5372..7194
 /note="previously sequenced by another group and can be
 found in GenBank Accession Number M63939"
 5587..6714
 /gene="tgt"
 5587..6714
 /gene="tgt"
 /codon_start=1
 /transl_table=1
 /product="transfer RNA-guanine transglycosylase"
 /protein_id="AAB40162.1"
 /db_xref="GI:1773090"
 /translation="MKFELDTDGRARRGLVFDRCGVETPCMPYCTGYTKMTP
 EVATQAQIILGNTFHLMLRPGQELMKLHGLDHDVQMGKPIITDSGGFQVPSLDIR
 KITEQGVHFRNPIDNPFLDPEKSMELGYDLSQIVMTFDECTPYPADMDYAKRME
 MSLRNAKRERERFDSGNKRALFGIIOGSVEEDLIDISYGLVDIGDGYAVGLANG
 EPKADHRIIEHYCPQIPADKPRYLMGVKPEDELVGVRGIDMPCVMPTRBAENGH
 LFTVDGVVTRAKKYSDDPDRDEGDCYTCRYSAYLHHLDRCNEIIGARLNTIHN
 LRYYQLMAGLRKAIEGKLESFVTFYQROGREVPLVND"
 complement (5644..5811)
 /note="previously sequenced by another group and can be
 found in GenBank Accession Number Z26340"
 6006..10444
 /note="previously sequenced by another group and can be
 found in GenBank Accession Number X56175"
 6426..7325
 /note="previously sequenced by another group and can be
 found in GenBank Accession Number S68715"
 6737..7069
 /gene="yajC"
 6737..7069
 /gene="yajC"
 /codon_start=1
 /transl_table=1
 /product="hypothetical 11.9 kd protein"
 /protein_id="AAB40163.1"
 /db_xref="GI:1773091"
 /translation="MSFRTSDAVATGAPAGSPMSLIMLVVGLIIFYFMILRPOOK
 RKREHKMLDSIAKDEVLITNGLVGRVTKVANGYIAIALNDTTEVLIKDFVAALV
 PKYTKMAL"

gene
 /note="previously sequenced by another group and can be
 found in GenBank Accession Number X56175"
 6426..7325
 /note="previously sequenced by another group and can be
 found in GenBank Accession Number S68715"
 6737..7069
 /gene="yajC"
 6737..7069
 /gene="yajC"
 /codon_start=1
 /transl_table=1
 /product="hypothetical 11.9 kd protein"
 /protein_id="AAB40163.1"
 /db_xref="GI:1773091"
 /translation="MSFRTSDAVATGAPAGSPMSLIMLVVGLIIFYFMILRPOOK
 RKREHKMLDSIAKDEVLITNGLVGRVTKVANGYIAIALNDTTEVLIKDFVAALV
 PKYTKMAL"

CDS

SOURCE Escherichia coli O157:H7
ORGANISM Escherichia coli O157:H7
REFERENCE Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
AUTHORS Makino, K., Yokoyama, K., Kubota, Y., Yutsudo, C.H., Kimura, S., Kurokawa, K., Ishii, K., Hattori, M., Tatsu, I., Abe, H., Iida, T., Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C. and Shinagawa, H.
TITLE Complete nucleotide sequence of the prophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7 derived from the Sakai outbreak
JOURNAL Genes Genet. Syst. 74 (5), 227-239 (1999)
MEDLINE 20198780
PUBMED 10734605
REFERENCE 2 (sites)
AUTHORS Ohnishi, M., Murata, T., Nakayama, K., Kuhara, S., Hattori, M., Kurokawa, K., Yasunaga, T., Yokoyama, K., Makino, K., Shinagawa, H. and Hayashi, T.
TITLE Comparative analysis of the whole set of rRNA operons between an enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an Escherichia coli K-12 strain MG1655
JOURNAL Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
MEDLINE 20557356
PUBMED 11108008
REFERENCE 3 (sites)
AUTHORS Yokoyama, K., Makino, K., Kubota, Y., Watanabe, M., Kimura, S., Yutsudo, C.H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T., Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C. and Shinagawa, H.
TITLE Complete nucleotide sequence of the prophage VT1-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli O157:H7 strain derived from the Sakai outbreak
JOURNAL Gene 258 (1-2), 127-139 (2000)
MEDLINE 20564182
PUBMED 11110150
REFERENCE 4 (sites)
AUTHORS Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C.-G., Ohtsuo, E., Nakayama, K., Murata, T., Tanaka, M., Tobe, T., Iida, T., Takami, H., Honda, T., Sasakawa, C., Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M. and Shinagawa, H.
TITLE Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12
JOURNAL DNA Res. 8 (1), 11-22 (2001)
MEDLINE 21156231
PUBMED 11258796
REFERENCE 5 (bases 1 to 295741)
AUTHORS Ohnishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shinagawa, H. and Hayashi, T.
TITLE Direct Submission
JOURNAL Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome Information Research Center, 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail:ken@gen-info.osaka-u.ac.jp, URL: http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365, Fax:81-6-6879-2047)
COMMENT genome project.
FEATURES
 source location/Qualifiers
 1. 295741
 /organism="Escherichia coli O157:H7"
 /mol_type="genomic DNA"
 /strain="O157:H7"
 /sub_strain="RIMD 0509952"
 /db_xref="taxon:83334"
 131..709
 /gene="EC60249"
 131..709
 /gene="EC60249"
 131..709
 /gene="EC60249"
 /note="similar to GMHA ECOLI gi|1786416 percent identity 100 in 192 aa (Conserved in E.coli K-12)."
 /codon_start=1
 /evidence=not experimental
 /transl_table=11
gene
 CDS
 /feature="phosphoheptose isomerase"
 /protein_id="BAB33672.1"
 /db_xref="GI:13359706"
 /translacion="MYODLRNLENAEFTLANFLKXDANIHAIQRAVILLADSPKAG
 GKVLSCNGSGSHCDMAHFAELTGRYENRPGYPALISGVSHISCGNPGFDIFS
 RVEVAGREBDVLGISTGNSANVIAKIAAAREKMKVITLTKDGKMAAGTADIEL
 RVHFGYADRIQIHLIKVHILILQLEKEMVK"
 814..1581
 /gene="EC60250"
 814..1581
 /gene="EC60250"
 /note="similar to YAFJ ECOLI gi|1786417 percent identity 99 in 255 aa (Conserved in E.coli K-12)."
 /codon_start=1
 /evidence=not experimental
 /transl_table=11
 /product="putative amidotransferase"
 /protein_id="BAB33673.1"
 /db_xref="GI:13359707"
 /translacion="MCEILGMSANVPTDICEFTGLVORGSGTGPHKDGNGITFEYEGK
 GCTFPDPOPSFNSPIAKLVQDYPIKSCSVAHIRQANRGEVLENTHPPTRELGRN
 MTVAHNGOLIGYSLETFGNRPVYKTDSEKAPCMILYKLTQRYPRPGMAAAYPKYA
 SLADELRQKGVFPMMLSDGRYVWYACSTNHHVITRRAAPGVAITLLDDVDIEDSSQTT
 PNDVVTIATQPLTGNETWOKIMPGEWRLFCLEERYV"
 complement (1552..2292)
 /gene="EC60251"
 complement (1552..2292)
 /note="similar to YAFK ECOLI gi|1786418 percent identity 100 in 246 aa (Conserved in E.coli K-12)."
 /codon_start=1
 /evidence=not experimental
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="BAB33674.1"
 /db_xref="GI:13359708"
 /translacion="MRKIALILMLLIPCVSFAGLLGSSSTTPSVKEYKQQLMGSPV
 YIQIFKEERLIDLVYMGEOYLIDSKYCKVSGGIGKRODFFKSPBGFVSQRNO
 LKEDSRVYKAINIGFPAAYRAHGYKIMIGDVCISPCYMTNGIDIEIFQFVYG
 ALVFGQPSVOVSTIYPPRMTDANKRHKYSFQKFMWQKRGVYFEGTRAPPTVSVN
 GRVYVYKPSHSEVYQQLASNTYTLPEAK"
 complement (2448..2708)
 /gene="EC60252"
 complement (2448..2708)
 /note="similar to YAFQ ECOLI gi|1786419 percent identity 98 in 84 aa but differs at N-ter. TTG start (Conserved in E.coli K-12)."
 /codon_start=1
 /evidence=not experimental
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="BAB33675.1"
 /db_xref="GI:13359709"
 /translacion="NNSGQFSKDVKLAQRHKMDKILYIMTLINNTLPLPAVYKDH
 PLGGSWKGYRDAVEPDDWILYKLTGKLRPFERTGHAALFG"
 complement (2727..2987)
 /gene="EC60253"
 complement (2727..2987)
 /note="similar to DINJ ECOLI gi|1786420 percent identity 97 in 86 aa (Conserved in E.coli K-12)."
 /codon_start=1
 /evidence=not experimental
 /transl_table=11
 /product="damage-inducible protein DinJ"
 /protein_id="BAB33676.1"
 /db_xref="GI:13359710"
 /translacion="MAANAFAVRARIDEDLKNOAADVLACGLTISDLVRLITKVAKE
 KALPFLRRPNQTLTOSIKNSGAVYVHAKKADDLFDKLGV"
 3197..3396
 /gene="EC60254"
 3197..3396
 CDS

```

/gene="ECs0254"
/note="similar to YAPL ECOLI g1|1786421 percent identity
98 in 249 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="putative lipoprotein"
/protein_id="BAB33677.1"
/db_xref="GI:13359711"
/translation="MSLSPIPSFVLSGLILICLPSPSPASATSHISPSYAROMON
RAILKQYQTLKQKQSYIVGNAESRAALRQNRQIKQPEPPAPLAKSDRMWA
LAENNHPLSSDHLNTEVALHRLQQLGKPYWGKTPDKFDCSGLVFAVAKTLE
AKPRTANEMVHYRATIVANNDRGDLFPFHHSREIADHMGVYLGQGFIESPT
GETIRVSRLEAPFWQDHLGARILTEETIL"
4122..4418
/gene="ECs0255"
4122..4418
/gene="ECs0255"
/note="similar to N-terminal part (1-98 in 165 aa) of
YAPL ECOLI g1|1786422 percent identity 93 in 98 aa"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="hypothetical protein"
/protein_id="BAB33678.1"
/db_xref="GI:13359712"
/translation="MSEYRRYIKGTMFTVLNRRSHLLTAQFQKLRAITNVR
DRPEINAMVVLPEHMCITWLPESDDFSSRWEIKKQFHAGCLKINMPR"
complement(4764..6503)
/gene="ECs0257"
complement(4764..6503)
/note="similar to FHR1 ECOLI g1|1786423 percent identity
98 in 578 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="flagellar biosynthesis"
/protein_id="BAB33679.1"
/db_xref="GI:13359713"
/translation="WLSRSDLLTLITNFIYVVGARISVSARFTLDMPGOMAI
DADNAGLINOAQOTRKDVASEDFGADGASKFVRGDAIAGMMLAINLIGVC
IGIRKYNLSADAAQOQVYMTIGDLVAQIPSLSLTAALIVRVSDNGIADHSH
QILASPSVLYATGATIMFYLAVPGMPLFLFSGALGTMRMSKQFAEAEKSL
ETLTRITETSEQWMTETLIPISLISGLKVALVDKQGPNGRIGVQVLS
DNGVILDEIRIRENFRKPSQVAFIPNGIADDAIDPADMLPSESTYGEIDGVL
GNDPAGMPTVWIGPAQKAKALNMGQYIDSAIYATIVNKVSYIFDLPNYDTITD
LHNRLSSMAPRLAEDLSALNYSQILKRYRALTEGVSARDIVITATVLVASSAVTD
HLLADVRLALRSITHPFVRKQELTVYTLNLEMLTNVVAQAQGGKMDSVF

```

Alignment Scores:

```

Pred. No.: 4,73e-34 Length: 295741
Score: 762.50 Matches: 168
Percent Similarity: 61.31% Conservative: 76
Best Local Similarity: 42.21% Mismatches: 129
Query Match: 37.77% Indels: 25
DB: 1 Gaps: 6

```

US-09-889-756a-2 (1-412) x AP002551 (1-295741)

```

QY 11 ALaAlaAlaLeuAlaAlaAlaValaAlaLeuValaLeuSerSerCyGlyLyGlyGlyasp 30
Db 270174 GCGGCGTTCGATGCTCTGACGACGCTTAGCCCTAACAGATG---GAGCACAACAG 270118
QY 31 ALaAlaGInGlyGlyGInProAlaGlyArgGluAlaProAlaProAlaValaGlyVala 50
Db 270117 GCCCAACAGGCGGCGAG-----CAGATGCGCCGCGGCTTGCGCTAGTA 270076
QY 51 ThrValHisProGInThrValaAlaLeuThrValaGluLeuProGlyArgLeuGluSerLeu 70
Db 270075 ACAAGTCAAACTGAACCTCTGACATCAACACCGAGCTTCGGGCTGCACCAAGTCCCTAC 270016
QY 71 ArgThrAlaAspValaArgAlaGlnValaGlyLeuIleGlnIlyArgLeuPheGlnGlu 90

```

```

Db 270015 CGGATCCGACAAGAGTTCCTCTCAAGTAGCGGAGATTATCTCGAAGCCTAAATTCAAAGAA 269956
QY 91 GlySerTyrrValaArgAlaGlnGInProLeuTyrrGlnIleAspSerTyrrTyrrGluAla 110
Db 269955 GGTAGCGACATCGAAGCAGGCTGCTCTCTATCAGTATATCTCGGACCTATCAGGCG 269896
QY 111 AsnLeuGluSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaTyrrAlaAsp 130
Db 269895 ACATAGCAGACGTGGAAAGGATCTGCGAAAGCCAGGCTGACGCAATATGCGCAA 269836
QY 131 AlaAspLeuAlaArgTyrrTyrrProLeuValaAlaAlaGluAlaValaSerArgGlnGluTyrr 150
Db 269835 TTGACGTGATCGTTATTCAGAAACCTGCTGATCACTAGTACATCACTAAGCAGAGTAC 269776
QY 151 AspAlaAlaValaThrAlaTyrrArgSerAlaGluAlaGlyValaTyrrAlaAlaAlaAla 170
Db 269775 GATCAGGCTCTGCTGATGCGCAACAGCGAATGCTGGTAACTGCGGCAAGAGCTGCC 269716
QY 171 IleTyrrSerAlaGlyIleAsnLeuAsnArgSerArgIleThrAlaProIleSerGlyPhe 190
Db 269715 GTTAAACTCGCGGATCATCTGCTTACACCAACATCACTCTCCGATTAGCGGTCCG 269656
QY 191 IleGlyGlnSerTyrrValaSerGlyGlyThrLeuLeuAsnAlaGlyAspThrThrValLeu 210
Db 269655 ATTGTATGATCGAAGCGTGAAGGCGCATTTGGTACAGAACGTCAGCGACTGCGCTG 269596
QY 211 AlaThrIleArgGlnThrAsnProMetTyrrValaAsnAlaThrGlnSerAlaSerGluVal 230
Db 269595 GCAACCGTCGACCACTTGATCCGATTCGATGATGACCCAGTCACACAGCAAGCTTC 269536
QY 231 MetTyrrLeuArgArgGlnIleAlaGlnGlyTyrrLeuLeuAlaAlaAspGlyValIleAla 250
Db 269535 CTGGCGCTTGAAGCAAGAACTGGGCAATGCGACGCTGAAGCAAGCAAGCAAGCAAGCA 269476
QY 251 ValGlyIleTyrrPheAspAspGlyThrValTyrrProGluTyrrGlyAlaGluLeuPheAla 270
Db 269475 GTGCGCTGATCAACAGTGAAGCGCATTAAGTCCCGAGAGCGGTACGCTGGAATTCCT 269416
QY 271 AspProValaValaAsnGluSerThrGlyGlnIleThrLeuArgAlaAlaValaProAsnAsp 290
Db 269415 GACCTTACCGTGTATGACACCACTGGGTCTATCACTTACCGCTATCTCCGAAACCG 269356
QY 291 GlnAsnIleLeuMetProGlyLeuTyrrValaArgValaLeuMetAspGlnValaAlaValaAsp 310
Db 269355 GATCACACTCGTCCGCGGATGTTGCTGCGGCAAGTCTGGAAGAAAGGCTTAATCCA 269296
QY 311 AsnAlaPheValaValaProGlnGlnAlaValaThrArgGlyAlaTyrrAsp-----ThrVal 328
Db 269295 AACCTATTTTATGTCGCGCAACAGGCGGTAAACCGTACCGCGCTGCGATGCCACCGTA 269236
QY 329 MetIleValaAsnAlaGlnGlyGlyMetGluProArgGluValaThrValaIleGlnGln 348
Db 269235 CTGGTGTGTGGCGGATGACCAAGTGAACCCGCTCCGATCGTTCGAAGCAGGCTATT 269176
QY 349 GlyThrAsnTyrrIleValaThrSerGlyLeuTyrrAspGlyAspTyrrValaValaGlnGly 368
Db 269175 GCGCATTAAGTGTGCTGCTGACAGAAAGTCTGAAGACAGGCAATGCTAGTAAAGTGG 269116
QY 369 IleSerIleAlaGlyIleThrGlyAlaAlaTyrrValaThrPro----- 382
Db 269115 CTG-----CAGAAAGTGGCTCTGCTGCTGCTCCAGTAAAGCA 269080
QY 383 LysGluTyrrPalasSerSerGluAsnGlnAlaAlaAlaProGlnSerGlyValaGln 400
Db 269079 CAAAGAGTTACCGCTGATTAATTAACCAAGCAAGCGCA-----AGCGGTGCTCAG 269032
RESULT 21
LOCUS AB016756 300817 bp DNA linear BCT 24-DEC-2002
DEFINITION Escherichia coli CT073 section 2 of 18 of the complete genome.
ACCESSION AB016756 AB014075
VERSION AB016756.1 GI:26106632

```



```

KEYWORDS      Escherichia coli CFT073
SOURCE        Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
ORGANISM      Escherichia coli CFT073
REFERENCE     1 (bases 1 to 300817)
AUTHORS       Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
              Raiko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
              Stroud,D., Mayhew,G.F., Rose,J., Zhou,S., Schwartz,D.C.,
              Petrá,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.
              Extensive Mosaic Structure Revealed by the Complete Genome Sequence
              of Uropathogenic Escherichia coli
              Proc. Natl. Acad. Sci. U.S.A. 99 (26), 17020-17024 (2002)
TITLE         2 (bases 1 to 300817)
JOURNAL       - Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
PUBMED        Raiko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
              Stroud,D., Mayhew,G.F., Rose,J., Zhou,S., Schwartz,D.C.,
              Petrá,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.
              Direct Submission
              Submitted (20-JUN-2002) Genetics Laboratory, University of
              Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES      Location/Qualifiers
source        1..300817
              /organism="Escherichia coli CFT073"
              /mol_type="genomic DNA"
              /strain="CFT073"
              /db_xref="taxon:199310"
              complement(294..1079)
              /locus_tag="c0330"
              complement(1294..1079)
              /locus_tag="c0330"
              /note="Residues 2 to 261 of 261 are 83.58 pct identical to
              residues 1 to 261 of 261 from GenPept.129 :
              >gb|AA122650.1| (AE008876) putative regulatory protein,
              deor family [Salmonella typhimurium LT2]"
              /codon_start=1
              /transl_table=1
              /product="putative deoxyribosome operon repressor"
              /protein_id="AAN78818.1"
              /db_xref="gi:26106633"
              /translation="MMETKOKERRRLRIELIKKTDRIHDKDAAMLEVSWVTPRDILH
              QEDPEPLTLIGGIYVWHKRAPMPICQVPRRHRRDLPAILAAGVNENDLIFPD
              NGDEIPLVISTIPDLPIFTGTGCSHRVFVALNEKPNTALICGGTYARSDAFYDASN
              SSSPLADISAPRKITFISAGVDHFGEVFNFPEDLATKKAKMARGLRTILRLHALFDVV
              ASASLPDISAPFVDVISDRPLDPDYTCRNASVXIVTDSSE"
              1382..2302
              /locus_tag="c0331"
              1382..2302
              /locus_tag="c0331"
              /note="Residues 1 to 306 of 306 are 96.07 pct identical to
              residues 1 to 306 of 306 from GenPept.129 :
              >emb|CAC34669.1| (AX0087925) unnamed protein product
              [Salmonella typhi]"
              /codon_start=1
              /transl_table=1
              /product="Putative ribokinase"
              /protein_id="AAN78819.1"
              /db_xref="gi:26106634"
              /translation="MDIATVIGSNVNDLIITYNQPKRGETLLPAEFLICGGCKCANQA
              VAAKLKSVLMLETKVDDIDPADFTINLSWSGINTTYVERVPTSGGVAIFPNANS
              SNSILTIKGANDKSLSPEDIDRAEMDLKKCLTVQLDVOLLETYHAIEFKQIGEVLT
              LNPALRELDMYSACKDCEFIPINETLEILTGSVDTYHIRLAASLVLDKGINNI I
              VTMSFGKALMWTROEVEHVAPKYNADVTSAGDAPFIGCRSHYYVGSDVEALKKKA
              LPFAFSVTYGKGTQSSTPSYSIEGFNFLTINE"
              2330..3646
              /locus_tag="c0332"
              2330..3646
              /locus_tag="c0332"
              /note="Residues 1 to 438 of 438 are 96.11 pct identical to
              residues 1 to 438 of 438 from GenPept.129 :
              >gb|AA122650.1| (AE008876) putative D-luciferase permease
              [Salmonella typhimurium LT2]"

```

gene
/codon_start=1
/transl_table=1
/product="Putative L-fucose permease"
/protein_id="AAW78820.1"
/db_xref="GI:26106635"
/translation="MNDKRIITPDGYLNTKPLPQFILLSCGLPFLMGCAALNIIILTLT
GRKVSLSLNSFASALVQSAFVGVYGLVLAIPASIVYTKTSKYVAILIGLTLIGCGTLT
PRASHMATYTNLFAITFAITGLSFLETPAANTYSMTQPKAYATLRNLISQTFPIGL
ASGILLPTCKVLEEGESLEKQMSGNNAQOIHNFKMLMENTLEPKYKIMLVVMMVIL
FLILTRPTCKVQATSHKRPSPAMDTLRILANPRVRRIQVQFLYGVQVWMSFTIR
LALLEGIDIERDASNPVYSEFACPIIGETFINILMTREPKVILYLSVIALFLAYV
ALAPSASVAYVILVSVLFGPCMATIVYGLTDYDNEHTENAGAVIWAIVGAAVVPR
TGGYIADMHSHISQLSFLVSMLCFVYGVYFPRRESKVRALAEVTS"
3658..4671
/locus_tag="c0333"
3658..4671
/locus_tag="c0333"
/note="Residues 1 to 336 of 337 from GenPept.129 :
>gb|AAW26649.1| (AE008876) putative cytoplasmic protein
[Salmonella typhimurium LT2]"
/codon_start=1
/transl_table=1
/product="Putative cytoplasmic protein"
/protein_id="AAW78821.1"
/db_xref="GI:26106636"
/translation="MSTRINLMRALFGKRPRIILNENDFTVTSFRYSGVEGLKIANS
RSHLLILPMCGQIMWDAQFDGHLGILTCMMFPQRPATEVIEITGCFAPHSGLLANCP
SLEDHLLHGEKACAMDEAWLELDGDLRLNGREYEVGHHYLAOPTVHLKST
LEDIDAVTNVLAISVIMPLQSCNNVAYIPNATPSQNPDEILRLRSPSHVNPQAO
WLAPNORIMQGEASLTISQPEFIDPELIVFPADKLIDATQDPERFRLSPDCTFTYTR
YSAELNVTYTRMLVNGEQVAAPALPATCPREGYLAARQNTLLIQVAPQOTRTFTVT
GIE"
complement(5594..6850)
/locus_tag="c0334"
complement(5594..6850)
/locus_tag="c0334"
/note="Residues 1 to 418 of 418 are 99.76 pct identical to
residues 1 to 418 of 418 from GenPept.129 :
>emb|CAC39291.1| (AJ278144) hypothetical protein
[Escherichia coli]"
/codon_start=1
/transl_table=1
/product="Putative integral membrane protein"
/protein_id="AAW78822.1"
/db_xref="GI:26106637"
/translation="MDFPRFLMSDVLSPRAVLVGLLALIGLIAKKPVEKCIQTYT
IMGFVILTAGKGLVYSIGDPFANIPOHAFGIQGVYNNRBAIVSAQKFSCKEMAMIF
FAMVNIINIAFTPWKPIFLTGHTLLEFSMMVAVILTAGTGTLLTAVSLVGVAM
VEFPPIAHVPKKYVSGSDVAIGHSTLSVLAQFISKPKNKSHSTEDMNVPSLSLFL
LADTVASLSPMSIILFLVTCLEFAGADAVKELSGCKMFMESIMOSITFAAGVYILOG
VMMVIAIVIPVAFKGISDGLVNNARPALDQPVFPYAPNAVILVGLSFAAGLIGMPVL
YLMNAIVIPCVVPHFPFGAAGVGNATGRRRAIIGAFQGLITFLVPLPVLVIG
DIGPANTTFSMDRFGALGILIGIVR"
complement(6863..7150)
/locus_tag="c0335"
complement(6863..7150)
/locus_tag="c0335"
/note="Residues 1 to 95 of 95 are 96.84 pct identical to
residues 1 to 95 from GenPept.129 : >gb|AAK27336.1|
(AF286670) unknown [Escherichia coli]"
/codon_start=1
/transl_table=1
/product="Hypothetical protein"
/protein_id="AAW78823.1"
/db_xref="GI:26106638"
/translation="MKTIVGCGNLGSLTLMEMSIKSLKDLISADVDHVLGSAKG
TPSDIYVGTKDIAOLVAQSVAGKIVALENNIDKKARERRLSVALTETGAL"
complement(7166..7609)
/locus_tag="c0336"
complement(7166..7609)
/locus_tag="c0336"
/function="Transport of small molecules"

ACCESSION	AE015072 AE005674
VERSION	AE015072.1 GI:24050630
KEYWORDS	
SOURCE	Shigella flexneri 2a str. 301
ORGANISM	Shigella flexneri 2a str. 301
REFERENCE	Bacteria: Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Shigella.
AUTHORS	1 (bases 1 to 11927) Jin, Q., Yuan, Z. H., Xu, J. G., Wang, Y., Shen, Y., Lu, W. C., Wang, J. H., Liu, H., Yang, F., Yang, J., Yang, F., Zhang, X. B., Zhang, J. Y., Yang, G. W., Wu, H. T., Dong, J., Sun, L. L., Xue, Y., Zhao, A. L., Gao, Y. S., Zhu, J. P., Kan, B., Chen, S. X., Yao, Z. J., He, B. K., Chen, R. S., Ma, D. L., Qian, B. Q., Wen, Y. M., Hou, Y. D. and Yu, J.
TITLE	Genome sequence of Shigella flexneri 2a: insights into pathogenicity through comparison with genomes of Escherichia coli K12 and O157
JOURNAL	Nucleic Acids Res. 30 (20), 4432-4441 (2002)
PIRMBD	12384590
REFERENCE	2 (bases 1 to 11927) Jin, Q., Shen, Y., Wang, J. H., Liu, H., Yang, J., Yang, F., Zhang, X. B., Zhang, J. Y., Yang, G. W., Wu, H. T., Dong, J., Sun, L. L., Xue, Y., Zhao, A. L., Gao, Y. S., Zhu, J. P., Chen, S. X., Yao, Z. J., Wang, Y., Lu, W. C., Qian, B. Q., Wen, Y. M. and Hou, Y. D.
AUTHORS	Direct Submission Submitted (08-MAY-2001) Microbial Genome Center of Chinese Ministry of Public Health, 100 Yingxin Jie, Xuanwu Qu, Beijing 100052, P.R. China
TITLE	Location/Qualifiers
JOURNAL	1. 11927
PIRMBD	/organism="Shigella flexneri 2a str. 301"
REFERENCE	/mol_type="genomic DNA"
AUTHORS	/strain="301"
TITLE	/serotype="2a"
JOURNAL	/db_xref="taxon:198214"
PIRMBD	131..703
REFERENCE	/gene="ybaY"
AUTHORS	/locus_tag="SF0398"
TITLE	131..703
JOURNAL	/locus_tag="SF0398"
PIRMBD	/note="Residues 1 to 190 of 190 are 100 pct identical to residues 1 to 190 of a 190 aa protein from Escherichia coli O157:H7 EDL933 ref: NP_286195.1"
REFERENCE	/codon_start=1
AUTHORS	/transl_table=11
TITLE	/product="glycoprotein-polysaccharide metabolism"
JOURNAL	/protein_id="AAN42054.1"
PIRMBD	/db_xref="GI:24050631"
REFERENCE	/translation="MKLVMAAGSLAVAIACADKSDSIQTPAPANTSISATQCPA IQPPNVSGTWIRQKVALPPDAVLVTLLSDSLADAPSKVLAQKAVREGQSPFV LPNPADVQPNARILLSAIVNDKLVFTDPTDVPVINGGTAKDLTLIVPQGNVAV QASGATTVTPSTSPQVNPSSAVPAPQY"
AUTHORS	796..2142
TITLE	/gene="y14I"
JOURNAL	/locus_tag="SF0399"
PIRMBD	796..2142
REFERENCE	/gene="y14I"
AUTHORS	/locus_tag="SF0399"
TITLE	/note="Residues 1 to 448 of 448 are 99 pct identical to residues 1 to 448 of a 448 aa protein from Shigella flexneri ref: NP_085177.1"
JOURNAL	/codon_start=1
PIRMBD	/transl_table=11
REFERENCE	/product="orf, conserved hypothetical protein"
AUTHORS	/protein_id="AAN42055.1"
TITLE	/db_xref="GI:24050632"
JOURNAL	/translation="MPPDFMAHIGQALDVSRYDSLBNPLTSLGDIYDLPBLSIKLAE SGVTLTKRRRLPLEMMVWCIVGMVLERKEPLHQTIVNRLDIMLPGRFPVPSAIVQAR ORGSAVAVRFTKTAOLAHNATPHPHWCGTLTADGVFRTPTDPTENAPAPROH AAGNPALVPOVKNVCOMELTSHLTAAAGAMKSENELABOLIBOTGDNLTLLMDKCY VSLGLNGSLAGGHRHWMIPLRKAQYBEIPRKIGKDHLYLKTSSQAKRPGLCN EYTRALLTRKGVCHLTSWTDMAKRPFGSMADLVSHWEILGYSRETIKQTMQSLR LTRSKRPPLVQELWGLAVNIVRYOMIKABHLKGYWPNOLSPSSCGWMMRIM
AUTHORS	TIQGSAGRIPELMEDLASMGQVLVLPTRRGRAPRVVKEKPMKXPTAPAKSSQSA" complement(2172..2582) /gene="ybaZ"
JOURNAL	/locus_tag="SF0400"
PIRMBD	complement(2172..2582)
REFERENCE	/gene="ybaZ"
AUTHORS	/locus_tag="SF0400"
TITLE	/note="Residues 8 to 136 of 136 are 100 pct identical to residues 1 to 129 of a 129 aa protein from Escherichia coli K12 ref: NP_414988.1"
JOURNAL	/codon_start=1
PIRMBD	/transl_table=11
REFERENCE	/product="orf, conserved hypothetical protein"
AUTHORS	/protein_id="AAN42056.1"
TITLE	/db_xref="GI:24050633"
JOURNAL	/translation="WQPIACOMLVSCAMRLHSGPDPYAEKLPQEEKKEKSDPPQV WQIVAAIPGVITTVGDVAKLAGSFRARQVGVKRLPEGSTLPVHRVNRHGTISL TSPDLQROKALLAGVWVSGSQIDILQRYWMY"
PIRMBD	2638..2751
REFERENCE	/gene="ffgB"
AUTHORS	2638..2751
TITLE	/gene="ffgB"
JOURNAL	/product="4.5S RNA; component of ribonucleoprotein particle"
PIRMBD	2862..3215
REFERENCE	/locus_tag="SF0401"
AUTHORS	2862..3215
TITLE	/gene="ybaA"
JOURNAL	/locus_tag="SF0401"
PIRMBD	/note="Residues 1 to 117 of 117 are 100 pct identical to residues 1 to 117 of a 117 aa protein from Escherichia coli O157:H7 EDL933 ref: NP_286197.1"
REFERENCE	/codon_start=1
AUTHORS	/transl_table=11
TITLE	/product="orf, conserved hypothetical protein"
JOURNAL	/protein_id="AAN42057.1"
PIRMBD	/db_xref="GI:24050634"
REFERENCE	/translation="MKYVDGFVAVVPADKQAVREMAKAPLRFKRGALRYECNAS DYPDGKVPDFMAVKAENREVRVSWILEPKKVRDANQKMSDPMKRGESMPD GKRMVYGGPESSTIDE"
AUTHORS	complement(3257..4813)
TITLE	/locus_tag="SF0402"
JOURNAL	complement(3257..4813)
PIRMBD	/locus_tag="SF0402"
REFERENCE	/note="Residues 1 to 518 of 518 are 99 pct identical to residues 1 to 518 of a 518 aa protein from Escherichia coli O157:H7 EDL933 ref: NP_286198.1"
AUTHORS	/codon_start=1
TITLE	/transl_table=11
JOURNAL	/product="orf, conserved hypothetical protein"
PIRMBD	/protein_id="AAN42058.1"
REFERENCE	/db_xref="GI:24050635"
AUTHORS	/translation="MLVYTRHLVGLISGVLLSVLPVGLSIMLAHQVETSEIED TYSSRVAIRANKVATQGDALQOELERQGAACSAHLMEMRVSYSRYQEOVYIDN NVPOCSSLSEHSPDPTPELGKISKQGYRWLTHSHNDLGIIRVWVWMTGAVVMDP ASFDIVPYSSWQIDAALIGANHVYTSDEIRAGIITTLQTPGSHIENNIYDI LPLEMINISITTWASTKMLQKGMKRVQVFWLPGLVIGLAAVFLKLRIDQSPHR LODALENNIDCVHQOPTVSLANGKIVGAELARWPQDGSMLSPDSFTPLAQOTGSE PLUTLLIIRSVFEDMGDCIRHPQOHSINIESVLTSEKIPOLAREMINHYQVNPQOI ALTELERFADPNTSAPILISYRGAHEIYLDGFTGYSLSYLDQDVIDLIKIDFS VDALEKRVTPHIIEMAKTLLKLVNVAAGISTSQEELVRQHGHYGCGMYKSLAPKE DFLRWABQHL"
JOURNAL	complement(4971..5480)
PIRMBD	/gene="y1aC"
REFERENCE	/locus_tag="SF0403"
AUTHORS	complement(4971..5480)
TITLE	/gene="y1aC"
JOURNAL	/locus_tag="SF0403"
PIRMBD	/note="Residues 1 to 169 of 169 are 99 pct identical to residues 1 to 169 of a 169 aa protein from Escherichia coli O157:H7 EDL933 ref: NP_286199.1"
REFERENCE	/codon_start=1

```

/transl_table=1
/product="orf, conserved hypothetical protein"
/protein_id="AAN42059.1"
/db_xref="GI:24050636"
/translation="MPEPKIQTIPRQMTETIORLTETIESIANREKNDKPREISF
IRKIPGFIQIVYAFPAFLVALVLOSEITSSGWLIVLFIILNGFFFDYPRRIED
IDVDFRCYNGEYNTREFPALVALLEINSPVADVHKEQLQKIVAKELSPDYLF
ILVARESTS"
complement(5557..6108)
/gene="y1ab"
/locus_tag="SF0404"
/pseudo
complement(5557..6108)
/gene="y1ab"
/locus_tag="SF0404"
/pseudo
complement(5557..6108)
/gene="y1ab"
/locus_tag="SF0404"
/notes="Similar to Escherichia coli K12 putative
transferase gi: 1786664 (184 aa). BLAST with identity of
98% in 183 aa. This CDS contains frameshift. The sequence
has been checked and is believed to be correct."
/pseudo
/codon_start=1
/transl_table=1
complement(6280..6510)
/gene="yha"
/locus_tag="SF0405"
complement(6280..6510)
/gene="yha"
/locus_tag="SF0405"
/notes="Residues 1 to 76 of 76 are 98 pct identical to
residues 41 to 116 of a 116 aa protein from Escherichia
coli p1r: S15270"
/codon_start=1
/transl_table=1
/product="yha protein"
/protein_id="AAN42060.1"
/db_xref="GI:24050637"
/translation="MGRMSERPLTKTDYMLRRCOTIDTLERVIEKNKXELSDNEL
AVFYSAADHRLAEILTMKLYDKIPSSVWKFR"
complement(6524..6898)
/gene="ybaJ"
/locus_tag="SF0406"
complement(6524..6898)
/gene="ybaJ"
/locus_tag="SF0406"
/notes="Residues 1 to 124 of 124 are 99 pct identical to
residues 1 to 124 of a 124 aa protein from Escherichia
coli O157:H7 EDL933 ref: NP_286202.1"
/codon_start=1
/transl_table=1
/product="orf, conserved hypothetical protein"
/protein_id="AAN42061.1"

```

Alignment Scores:

Pred. No.:	Length:	Score:	Matches:	Conservative:	Mismatches:	Best Local Similarity:	Query Match:
1.08e-35	11927	759.50	168	76	129	42.21%	37.62%
Percent Similarity:		61.31%					
Best Local Similarity:		42.21%					
Query Match:		37.62%					
DB:							
US-09-889-756a-2 (1-412) x AE015072 (1-11927)							

US-09-889-756a-2 (1-412) x AE015072 (1-11927)

```

Qy 71 ArgThrAlaaspValArgAlaGlnValGlyGlyIleIleGlnIlysarGleupheGlnGlu 90
Db 11620 CGAGTCCAGAGAGTGTCTCTCAAGTTAGCGGGATTATCTCGAAGCTAATTCAGAA 11561
Qy 91 GlseryrValArgAlaGlnProLeuTyGlnIleAspserThrTyGlnAla 110
Db 11560 GGTAGCCAGATCGAAGCTGTGCTCTCTATCAGATTATCTCGACCTTCACAGCG 11501
Qy 111 AsnleuGluSerAlaArgAlaGlnIleuAlaThraAlaGlnAlaThraAlaValAsp 130
Db 11500 ACATACGACAGTGGGAAAGGTATCTGGCAAAAGCCAGGCTGCAGCAATATCGCCAA 11441
Qy 131 AlaaspleuAlaArgTyLysProLeuValAlaAlaGlnAlaValAspArgGlnTyLys 150
Db 11440 TTGACGTGAATGCTTATCAGAAAGTCTCGTACTAGTACATCATAGTAAGCAAGGTAC 11381
Qy 151 AspAlaAlaValThrAlaLysArgSerAlaGlnAlaGlyValLysAlaAlaGlnAla 170
Db 11380 GATCAGGCTCTGGCTGATGCGCAACAGGCAAGTCTCGGTTAATCTCGCGCAAGCTGCC 11321
Qy 171 IleLysSerAlaGlyIleAsnleuAsnArgSerArgIleThraLaprolleSerGlyPhe 190
Db 11320 GTTGAAACTGCGGGATCAATCTGGCTTACACCAAGTCACTCTCGATTAGCGGTCGC 11261
Qy 191 IleGlyGlnSerLysValSerGlnGlyThrLeuLeuAsnAlaGlyAspThrThrValLeu 210
Db 11260 ATTGGTACGTGCAACGACGAGGAAGCGCATTTGTCACAGACGTCAGGAGCTGTCTG 11201
Qy 211 AlaThrIleArgGlnThrAsnProMetTyValAsnValThrGlnSerAlaSerGlnVal 230
Db 11200 GCAACCGTGACGCAACTGTATCCGATCTGATGTGATGACCAAGTCCACCAACGACTTC 11141
Qy 231 MetLysLeuArgArgGlnIleAlaGlnGlyLysLeuLeuAlaAlaAspGlyValIleAla 250
Db 11140 CTTCGCTGAAACGAGCACTGGCAATGCAAGCTGAAACAGAGCAAGCAAGCAAA 11081
Qy 251 ValGlyIleLysPheAspAspGlyThrValTyProGluLysGlyArgLeuLeuPheAla 270
Db 11080 GTTCGCTGATCAACCACTGACGATTAAGTTCCTCGGACGAGACGTTAGCTGAATTCT 11021
Qy 271 AspProValValAsnGlnSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAsp 290
Db 11020 GACCTTACCTGTGATCAGACCACTGGGCTATATCACTTACGCGCTATCTCCGAAACCG 10961
Qy 291 GlnAsnIleLeuMetProGlyLeuTyValArgValLeuMetAspGlnAlaAlaValAsp 310
Db 10960 GATCACACTCTGCTGCGGGATGTTGTGCGCGCACGCTGGAAGAGGCTTATCA 10901
Qy 311 AsnAlaPheValValProGlnGlnAlaValThrArgGlyAlaLysAsp-----ThrVal 328
Db 10900 AACGCTATTTTAGTCCCGCAACAGAGGCGTAAACCGTACGCCGCGTGTATGCCACCGTA 10841
Qy 329 MetIleValAsnAlaGlnGlyMetGlnProArgGlnValThrValAlaGlnGlnGln 348
Db 10840 CTGGTAGTGGCGCGGATGACAAAGTGAAACCGTCCGATCGTTGCAAGCAGCGATATT 10781
Qy 349 GlyThrAsnTrpIleValIleThrSerGlyLeuLysAspGlyLysPheValValGlnGly 368
Db 10780 GGCATTAAGTGGCGTGACAGAAAGGCTGAAAGCAGGCGCATCGCTAGTAATTAAGTGG 10721
Qy 369 IleSerIleAlaGlyIleThrGlyAlaLysLysValThrPro----- 382
Db 10720 CTG-----CAGAAAGTGGCTCTGTTGTCACAGTTAAAGCA 10685
Qy 383 LysGlnTrpAlaSerSerGluAsnGlnAlaAlaProGlnSerGlyValGln 400
Db 10684 CAAGAATTACCGCTATTAATTAACCAAGCAAGCGCA-----AGCGTGTCTCAG 10637

```

RESULT 23

AE016979/c 292309 bp DNA linear BCT 22-APR-2003

DEFINITION Shigella flexneri 2a str. 2457T section 2 of 16 of the complete

genome.
 accession AE016979 AE014073
 version AE016979.1 GI:30040081
 keywords
 source
 organism
 Shigella flexneri 2a str. 2457T
 Shigella flexneri 2a str. 2457T
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Shigella.
 reference
 1 (bases 1 to 292309)
 Wei, J., Goldberg, M.B., Burland, V., Venkatesan, M.M., Deng, W.,
 Fournier, G., Mayhew, G.F., Plunkett, G. III, Rose, D.J., Darling, A.,
 Mau, B., Perna, N.T., Payne, S.M., Runyen-Janecky, L.J., Zhou, S.,
 Schwartz, D.C. and Blattner, F.R.
 Complete Genome Sequence and Comparative Genomics of Shigella
 flexneri Serotype 2a Strain 2457T
 Infect. Immun. 71 (5), 2775-2786 (2003)
 12704152
 2 (bases 1 to 292309)
 Wei, J., Goldberg, M.B., Burland, V., Venkatesan, M.M., Deng, W.,
 Fournier, G., Mayhew, G.F., Plunkett, G. III, Rose, D.J., Darling, A.,
 Mau, B., Perna, N.T., Payne, S.M., Runyen-Janecky, L.J., Zhou, S.,
 Schwartz, D.C. and Blattner, F.R.
 Direct Submission
 Submitted (13-JUN-2002) Genetics Laboratory, University of
 Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
 features
 source
 1..292309
 location/Qualifiers
 /organism="Shigella flexneri 2a str. 2457T"
 /mol_type="genomic DNA"
 /strain="2457T"
 /serotype="2a"
 /db_xref="taxon:198215"
 complement(165..935)
 /gene="yafY"
 /locus_tag="S0290"
 complement(165..935)
 /gene="yafY"
 /locus_tag="S0290"
 complement(165..935)
 /function="putative enzyme; Not classified"
 /note="residues 1 to 256 of 256 are 98.04 pct identical to
 residues 1 to 256 of 256 from Escherichia coli K-12 :
 B0219"
 /codon_start=1
 /transl_table=11
 /product="putative RC 3.5. amidase-type enzyme"
 /protein_id="AAPI5815.1"
 /db_xref="GI:30040082"
 /translation="MPSLKITLLQOPLVWVDGPNLHFDROLGRTGDIYVLEMF
 TSGFAMEAASSLAODDVVWMTAKAQQCNALLAGVALQTESGVRFLLVPGGIV
 HFVDRHLFPMADDEHLHYKAGNARVIVEMGMRLLPLVCYDLRFPVSRILNDYDAL
 YVANWPAPRSILHMOALLTARAENQAVACNRYGSDGNGCHRGDSRVINPGGEIILA
 TAAHKATRIDAEISWALREYRKRFPMWDADEFRLM"
 1089..1562
 /gene="yafE"
 /locus_tag="S0291"
 1089..1562
 /gene="yafE"
 /locus_tag="S0291"
 /note="residues 1 to 157 of 157 are 90.44 pct identical to
 residues 1 to 157 of 157 from Escherichia coli K-12 :
 B0220"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAPI5816.1"
 /db_xref="GI:30040083"
 /translation="MGRISSGMMFKAITVAALVITASMAODDITISLAKETTK
 AAFNOMVQGHKLPAWVMKGGTTPACVTGLTDEGVMSACKPHDCSOSIAVWMSPK
 SNOMTGLFSAIDETSOEKLTMNVNDALSDGKTIVFALITSLNHPDGRFK"
 complement(1605..4085)
 /gene="yafH"
 /locus_tag="S0292"
 complement(1605..4085)

/gene="yafH"
 /locus_tag="S0292"
 /function="putative enzyme; Not classified"
 /note="residues 1 to 826 of 826 are 97.82 pct identical to
 residues 1 to 826 of 826 from Escherichia coli K-12 :
 B0221"
 /codon_start=1
 /transl_table=11
 /product="putative acyl-CoA dehydrogenase"
 /protein_id="AAPI5817.1"
 /db_xref="GI:30040084"
 /translation="MVRPPTSKGLFVMMIISITATVVLGALFYHRVSLFISLILIA
 WTALGVAGLMSAMVLPVLTIIIVPFAPARSKSMISAPFRGRKLPMPMSREKKA
 IDPMTWEDLFOGKMDKLNHYPOPLTAEOAPLDGPVEACAMNDPQITLSEA
 ADLPELMVLAKEHAPFPAMIKKEGYGLSEAAQSVLVOLKSGVSTIIITGVPS
 LGPELLOHYGTDEOKDHYLPRLRGOETCPPLTSPSAGSDAIPDGTIVMGEO
 GQOVLGMRITWNRKTYITLPIATVYLAFSLSPKILGCAEDLGTICALIPTTPPV
 EIGRRHPLNVPQNGPTRGKDVFPVLDYIIIGKPKAGQGMRLVECLVSGRGLTPS
 NSTGKVSVALAGAYSHIRQPKISIGKEGIEEPLARLAGNVAADAAASLITVGI
 MLEKPAVLSATVRYHCTHRGQOSIIDAMDITGCGIMLGOSNLAAYOGAPLITV
 EGANILTRSMWIFGQGAIRCHPYVLEMEAANKNDVAPFKLFEKHIGHGNSKVRSE
 WLGITRGITSTSTPGATKRYOHNRSLNALLSVSMALLGSGSLKREPRISARG
 DILSQTLYASAVLKRIIDEGRNADEPLVHGCQDALYQAEQAMNDLQNPNNRVAG
 LNVVIFPTGRHYLAPSDKLDHMAKILQVNPATRSRIIGSGOYLTPSEHNPGLLEA
 LVNVIADPPIHQRIKCEIKNLPEPTRIDELAHNALTKGLIDKDEAALIVAEBSRLRS
 INVDDPPEELATKPVLPKVRKVEAA"
 4289..4867
 /gene="gmbA"
 /locus_tag="S0293"
 4289..4867
 /gene="gmbA"
 /locus_tag="S0293"
 /function="enzyme; Surface polysaccharides"
 /note="residues 1 to 192 of 192 are 100.00 pct identical
 to residues 1 to 192 of 192 from Escherichia coli K-12 :
 B0222"
 /codon_start=1
 /transl_table=11
 /product="phosphohexose isomerase"
 /protein_id="AAPI5818.1"
 /db_xref="GI:30040085"
 /translation="MYQDLIRNELNEAETLANFLKDDANIHAIQRAAVLADSPKAG
 GKVLSCGNGSHCDAMHFABELTGRYRENPQYPAIALSPVSHISCGNDGPNDFIS
 RYVEAVGRGDVLLGISGNSANVKAIAAAREKMKVITLTKXGKGKAGAGADIEI
 RVPHFGYADRIQGIHKKVHIILQILEKMKV"
 4875..5009
 /note="RBP (repetitive extragenic palindromic) element,
 contains 3 RBP sequences"
 5073..5840
 /gene="yafJ"
 /locus_tag="S0294"
 5073..5840
 /gene="yafJ"
 /locus_tag="S0294"
 /function="putative enzyme; Not classified"
 /note="residues 1 to 255 of 255 are 99.60 pct identical to
 residues 1 to 255 of 255 from Escherichia coli K-12 :
 B0223"
 /codon_start=1
 /transl_table=11
 /product="putative amidotransferase"
 /protein_id="AAPI5819.1"
 /db_xref="GI:30040086"
 /translation="KEELIGMSANVPTDTCFSTGLVQNGGCTGPKHCKGKITFYEGK
 GGRTFNDQPSFNSPIAKUVQDPIKSCSVVAHITROANGEVALENTHPTRELWERN
 KTYAHNGQLTGYKSLGTGNFRPVGETDSKAFQWLLHKLTORVPRTPRMAAAVFKYIA
 SLADELRAGVFNMLISIDGRYVAYCSTYIHHITRRAPGVATLLDQVEIDPSQOTT
 PNDVAVITOPLTGNETMKYKIMPGWMRIFCCGERVAV"
 complement(5811..6551)
 /gene="yafK"
 /locus_tag="S0295"
 complement(5811..6551)
 /gene="yafK"

DEFINITION Proteus mirabilis regulatory protein AcrR, membrane fusion protein AcrA, and inner membrane RND family protein AcrB genes, complete cds.

ACCESSION AY061647

VERSION AY061647.1

KEYWORDS GI:17973457

SOURCE Proteus mirabilis

ORGANISM Proteus mirabilis
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Proteus.

REFERENCE 1 (bases 1 to 5281)
Ysallli, M.A., Murphy, E., Projan, S.J. and Bradford, P.A. AcrAB multidrug efflux pump is associated with reduced levels of susceptibility to tetracycline (GAR-936) in Proteus mirabilis Antimicrob. Agents Chemother. 47 (2), 665-669 (2003)

JOURNAL MEDLINE 22430550

PUBMED 12543675

REFERENCE 2 (bases 1 to 5281)
Ysallli, M.A., Murphy, E., Projan, S.J. and Bradford, P.A. Direct Submission
Submitted (01-NOV-2001) Infectious Disease, Wyeth Ayerst Research, 401 N. Middletown Rd., Pearl River, NY 10965, USA

JOURNAL TITLE Location/Qualifiers

FEATURES

Source 1..5281

1. 5281

/organism="Proteus mirabilis"

/mol_type="genomic DNA"

/db_xref="taxon:584"

complement (140..784)

/codon_start=1

/transl_table=1

/product="regulatory protein AcrR"

/protein_id="AAL32124.1"

/db_xref="GI:17973458"

/translation="MARKTRROQETROQIIDALRLPTVQGSATSLDIAEAGT
RGALVHFKKVDLPTACELTDKISLEBESKIPDPLVRLLYLITLSTVE
DPKHALLETYFKCEFGVEMTPIVEIRLRCADYRISQISRCIEKQALDPL
RRALMLRAMTGLAENWLPSPESFISKEOYLVDSPIDMKHVMRISALS"

910..2100

/codon_start=1

/transl_table=1

/product="membrane fusion protein AcrA"

/protein_id="AAL32125.1"

/db_xref="GI:17973459"

/translation="MRKRGVLPALVLVLSGLVAVGCGDKNOSAGGPPAPAVVT
LDKPLITTDLPGRSARVIAEYRPOVGGIILKRNTEPESYAGSLVQIDPAIFQ
ATNSAOADAKAKAEIARLVERKPLGLTNVYSKODPRTATSGYAAVAAKRA
EATVNTAKINLETKVTAPISGSGASTVEGALVAVGQVVALTVQOQIDPIVDTQ
SSSDYKLKNEISGLIRQEGKRVVHLITNSQYAQKGLFEFSDTVVDESGSITM
RAIVNPKGELPQGFVETRLNGIRONALILPOQAVIRPREGATTVVKNKVVVEV
RTIEVQAVGNKWLNVSGVQDVRVIVSGDKQAPKPKVTPORENIDATSTKSEBA
KDHQ"

2115..5273

/codon_start=1

/transl_table=1

/product="inner membrane RND family protein AcrB"

/protein_id="AAL32126.1"

/db_xref="GI:17973460"

/translation="MKRFIDRPVAVITATMAGLALILKLPVAVPTTAPPAIS
ISAVPGADATVQNTVQVIEQNMNDIMVWSTSDAGMNTITLFEATDIDI
AQVOVOKRLAMPLEQEVQOQISVDKSSSFLVAVAGISQDSGSDIDVYGA
TIDQPLSRVGETQLFTQVAMRIWLDKLVKYMTLVYNIAIKSNQNVAVAG
LGSGPVPGORLNVSIITQRTNPEQFALIMKNDGQVQLKDLGTYDEMYEVS
TIRFNGLPAGGIGIKATGANALDTRARALADMEPFGSLVAVYVPTPVK
ISIFEVKTIVEAIVLVITVIFLQNFRTITPTIIVPVVLGTRAITISAFISINT
LTFPAVLAIGLVDAIVVENVERMOBESPEKATRKSGQIGALVAGLVALS
AVEIPMAFEGSGTALVRSVTVISAMVLSVALILPALCATLKPSPKSHGO
TGEFFGNVRFESKSHYTDVSGRTLKGRTYLLIYVLAAMAIFVRLPAPLEEB
DQGVLLTVOLPAGSTOEBOTDVLQVNDVFNTEKDVVSVFTVSGFSGGQGGON
LATVAVNDEKESDQDKVPAIVGRANOLSOIKKAVFVSFNIPTATVSGSAGPFE
LIDNANIGHEKIMEARQNLGMAQGBQMLTVRPNQEBTSQRLITDIEKQAQGV
AVSDITSLTCTMGSTVNDIFIRGRKAKTVAGSOFRLPDIDANLYRNNAGMV
PFSSFTDTSKDPWLKSGPRLRYNGLPAVQIQSATPGSSGAMLMEDLASKPSG
IGYEWGMSYOEIRLSGNQAPALYISLIVFLCLALALYESWSVPMVLVPLVIGVA

CDS

BASE COUNT 1497 a 1065 c 1134 g 1585 t

ORIGIN

Alignment Scores:

Aligned No.: 4,25e-36 Length: 5281

Score: 758.50 Matches: 170

Percent Similarity: 60.71% Conservative: 71

Best Local Similarity: 42.82% Mismatches: 139

Query Match: 37.57% Indels: 17

DB: 1 Gaps: 6

US-09-889-756a-2 (1-412) x AY061647 (1-5281)

QY 12 AAlaAlaLeuAlaAlaValAlaLeuValLeuSerSerCysGlyIlysglyIyAspAla 31

Db 940 GCTCTGTAGTGGCTTTCAGGACGCTGTTGTGTGATGTGGCGACAAATTCAGTCT 999

QY 32 AAGlGlyIyGlyIyGlyIyGlyIyGlyIyGlyIyGlyIyGlyIyGlyIyGlyIy 51

Db 1000 GCT---GGAGGTCACCT-----CCTGCTCTCTCTGTAGGTTGTGTACA 1041

QY 52 ValHisProGlnThrValAlaLeuThrValGluLeuProGlyIyArgLeuGlySerLeuArg 71

Db 1042 TTGAGTCGGAACACGACTATCACAGACACTTCTGCTGATCATCTGCTTACCT 1101

QY 72 ThrAlaAspValArgAlaGlnValGlyIyIleIleGlnIlyAspLeuPheGlnGlyIy 91

Db 1102 ATGCGAGAGGTGGCTCTGAGTTGGCGGCATCATTTAAACGCAATTACCGAAGGT 1161

QY 92 SerTyrValArgAlaGlyIyGlyIyGlyIyGlyIyGlyIyGlyIyGlyIyGlyIy 111

Db 1162 AGTTATGTAAACACGAGACATCTTTATACCAATGCATCTCTATTTTCAAGTACA 1221

QY 112 LeuGlySerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaIlyAspAla 131

Db 1222 TTAAACAGTCTCAAGCTATTTAGCAAAAGCAAAAGCAATCTGAATTCCTGCTG 1281

QY 132 AspLeuAlaArgTyrIlySerProLeuValAlaAlaGlnAlaIlySerArgGlnIyTyrAsp 151

Db 1282 ACTGTGAGCGGTATTAACCTCTACTCGGACCAATTAATGTCAAGAAATTTTGAT 1341

QY 152 AlaAlaValThrAlaIlyAspSerAlaGlnAlaGlyIyValIlyValAlaAlaIle 171

Db 1342 ACCGCAACATCTGATGACGCTCAAGCTGTGTCGAGTAAACCACTGAAGCTACAGT 1401

QY 172 LysSerAlaGlyIlyLeuLeuAsnAlaGlySerArgIleThrAlaProIlySerGlyPheIle 191

Db 1402 ACTAATGCAAAATTTAATCTTGAATATACCAAGTACCCGACCAATTTCTGCGCTTCA 1461

QY 192 GlyIlySerIlyValSerGlyIyThrLeuLeuAsnAlaGlyAspThrThrValLeuAla 211

Db 1462 GGTAAATCAACGTTACAGAGGCGCATTAAGTGGCCCTGTCAGCAAGGCGTTTAAACA 1521

QY 212 ThrIlyArgGlnThrAsnProMetTyrValAsnValThrGlnSerIlyAspGlyValMet 231

Db 1522 ACTGTTCACCAATTCATCGATTTATGTGCAAGTACACCAATCGATGAAGATTATTC 1581

QY 232 LysLeuArgArgGlnIleAlaGlnGlyIlyLeuLeuAlaAlaAspGlyValIleAlaVal 251

Db 1582 AAATTTAAATAATTAATTAAGAGTGCATTTATTCGCAAGAGCAAGCAAACTGTTGTT 1641

QY 252 GlyIlyLeuPheAspAspGlyThrValTyrProGlyIyGlyIyArgLeuLeuPheAlaAsp 271

Db 1642 CACTTAACCTTAATTAATGATGCAATCTCAAGCTCAAAAGGTCACTTAAGATTCTCAT 1701

QY 272 ProValIlyAsnGlySerThrGlyIyGlnIleThrLeuArgAlaAlaValProAsnAspGln 291

Db 1702 GTGACGTTGATTAATCCACAGGCTCTATCATTGCGTGCATCTGATCTTAATCTCTAA 1761

QY 292 AsnIleLeuMetProGlyLeuTyrValArgValLeuMetAspGlnValAlaValAspAsn 311

LOCUS	DEFINITION	ACCESSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT	FEATURES	source	
Db	1762	CGTGAATTACTGCGCTGGTATGTTGTTGTCACAAATTAAGAAAATGCTATTCGTCAAAAT	1821														
Qy	312	Ala pheValValProGlnGlnAlaValThrArgGlyAlaLysAsp-----ThrValMet	329														
Db	1822	GCTATTTTTCATCCCAACAAAGCTGTTATTTGTTACTCCTCGTGGTGAAGCAACAACATG	1881														
Qy	330	IleValaAsnAlaGlnGlyGlyMetGluProArgGluValThrValAlaGlnGlnGly	349														
Db	1882	GTCGTTAATAAAGAACACACAGCTGGTGAAGTGGTCTACTGAAGATATCAACAGCAGTTGGT	1941														
Qy	350	ThrAsnTPPIleValThrSerGlyLysLeuLysAspGlyAspLysValValGluGlyIle	369														
Db	1942	AATAAGTGGCTAGGAATAGTGGTGTCAAGTTGGTGTGACCGCGTATGTTCCGGGCTTA	2001														
Qy	370	SerIleValGlyIleThrGlyValAlaLysLysValThrProLysGlu-----	384														
Db	2002	CAAAAGCA-----CAACGTGAATATGAAGGTATGCGCCACAAGAAAGAAATTTATGATCA	2055														
Qy	385	TrpAlaSerSerGluAsnGlnAlaAlaAla-----ProGlnSerGlyVal	399														
Db	2056	ACAGCATCACTGAAAAATCTGAACACGACGAGATGCTCATATAGAGATC	2106														
RESULT 25	AL646081/c	197050 bp	DNA	linear	BCT 11-APR-2003												
LOCUS	Al646081	197050 bp	DNA	linear	BCT 11-APR-2003												
DEFINITION	Ralstonia solanacearum GM11000 megaplasmid, complete sequence;																
ACCESSION	Al646081	Al646053															
KEYWORDS	Al646081.1	GI:17431274															
SOURCE	Ralstonia solanacearum																
ORGANISM	Ralstonia solanacearum																
REFERENCE	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;																
AUTHORS	Burkholderiaceae; Ralstonia.																
	1	Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S.,															
		Artiat, M., Billault, A., Brottier, P., Camus, J. C., Catolico, L.,															
		Chandler, M., Choisme, N., Claudel-Renard, C., Cunha, S., Demange, N.,															
		Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, M., Schlex, T.,															
		Siquier, P., Thebaud, P., Whalen, M., Wincker, P., Levy, M.,															
		Weissenbach, J. and Boucher, C. A.															
		Genome sequence of the plant pathogen Ralstonia solanacearum															
		Nature 415 (6871), 497-502 (2002)															
		2 (bases 1 to 197050															

/function="miscellaneous: hypothetical/partial homology"
 /note="Product confidence : probable
 Gene name confidence : hypothetical
 predicted by Codon usage
 predicted by Homology
 predicted by Framed"
 /codon_start=1
 /evidence=not_experimental
 /transl_table=1
 /product="PROBABLE SIGNAL PEPTIDE PROTEIN"
 /protein_id="CAD17954.1"
 /db_xref="GI:17431275"
 /db_xref="GOA:O8XRM8"
 /db_xref="SPTRMBL:O8XRM8"
 /translation="MKTSFKUTGTALAVATFAFVLTISASAASALDPLPIVAVGSSFL
 AGCVASITVANGLPPTTISIGATPASSSEVETTLPGALGTDTLVAVVGSVNTGPS
 SAVSAGVSTVNVPTVASTAGGALSGIGGGNALASVQGTGVGVVSGSDPSG
 ALSGASVGTALGGSGNATGALGVGTGALGNNPPTGALNGVIGTATGALGVNS
 GSDPSALSGVNTVTGTLGNSPAGLNVGTGTGALGVNSGSDPSALSGVNTV
 TGTGNSPALANGVGTGTGALGVNSGSDPTGALSGVNTVTGTLGNSPTGALNG
 VGTGTGALVNAAGSDPTGALSGVNTVTGTLGNSPAGLNVGTGTGALGVNSGSD
 DPTGALSGVNTVTGTLGNSPTGALNGVGTGTGALGVNSGASNLAPVQGVTVGV
 STLSSAGSSPTTITSLVGTIONALPGVGNPAGALTGAVGTGALGVNSGSGALP
 VGVNVGVNTLGGSNPTGALGVNSGTGALGGSNPTGALGVNSIGTIGAPVSSVPS
 LTPSGGSGVGVGVGVAGVAGVAGSGLIGSVNVNVSTVSNAGVGVVNAVGNAA
 AGTVATVPGVSVSTTNSNGTTTIOPLAPVTSLTNLTGALPR"
 complement(816.1494)
 /note="RSJ7 or RS06124
 LONG REPEAT 37 +
 predicted by Homology"
 /evidence=not_experimental
 complement(943.1611)
 /note="RSJ7 or RS06125
 LONG REPEAT 37 +
 predicted by Homology"
 /evidence=not_experimental
 complement(2852.4954)
 /gene="fusa2"
 /note="synonyms: RS01903, RSP0804"
 complement(2852.4954)
 /gene="fusa2"
 /function="macromolecule metabolism; macromolecule
 synthesis, modification; proteins and peptides -
 translation and modification"
 /note="Product confidence : probable
 Gene name confidence : probable
 predicted by Codon usage
 predicted by Homology
 predicted by Framed"
 /codon_start=1
 /evidence=not_experimental
 /transl_table=1
 /product="PROBABLE ELONGATION FACTOR G (EF-G PROTEIN)"
 /protein_id="CAD17955.1"
 /db_xref="GI:17431276"
 /db_xref="GOA:O8XRM7"
 /db_xref="SWISS-PROT:O8XRM7"
 /translation="MPROTPIERYRNIGISAHIDAGKTTTTERILFYTVNKHIGIEVH
 MGAAITMDMEOEERGITITSAATHCMWKGNGYPHRINIIDTPEHVFTEVERKS
 DRVALGAMVDSVGVPOSETVWRQANKYKPIAIVNNMDRVGDAFVRVEKMS
 RLKGNPVVQPVGAEDHFRGVNLTMKALVMDASQGVKFEYIDIPELRLPAAQW
 HDKNVAALDEADALEKTLGGEALTEAEIKGALRRRTTIGETIYPMLCGSAFRKQGV
 ALMDVAVDEGSPIDIPETIQGHGKDEPFLERHANDPFSALFKIMTDPVQQLTFP
 RYVSSGVSGGVTVNVPVKEKERLGRILIQMANRVEIKVHAGDIAAAGLKAATGG
 DTLCPDNVILITERNEPEPEVISOAPEPTKGDEKMGALNRLAODPSFPRVTDDEE
 SGQITISMGELHILVDPMKREFGVEATYGRQVAVYRETIKRVADVDEGKFLKQSG
 GCGTGVHATVLDQPNPKGTFEVDIEKGVIPREFIPAVDKGIRDTIANTAGYAPVIV
 DVKRLTGVSHVDVNSENARMGSAFQKQMRAPILDEPMNAEVEIPETVGNAG
 VIGDLSRRRGVQGTEDI PGGGGKVVAHEVPLAEMFGYSTNLRSLSGRAITYTEFKH
 YAEAPRTVEVTSIAKRA"
 complement(5078.5797)
 /gene="RSB0805"
 /note="synonyms: RS01902"

CDS
 complement(5078..5797)
 /gene="RSP0805"
 /function="miscellaneous; not classified regulator"
 /note="Product confidence : putative
 Gene name confidence : hypothetical
 predicted by Codon_usage
 predicted by Homology
 predicted by Framed"
 /codon_start=1
 /evidence=not_experimental
 /transl_table=1
 /product="PUTATIVE TRANSCRIPTION REGULATOR PROTEIN"
 /protein_id="CAD17956.1"
 /db_xref="GI:17431277"
 /db_xref="GOA:O8XRM6"
 /db_xref="SPTREMBL:O8XRM6"
 /translation="MDLAPSPALPGAAVPERHASSWEGALCTPIQDALIGMAVVR
 RLGGDMLEFRGDPDGLYCVVSGAIGRIGATSAGRESLLAVEPVMFGELVDRQ
 ARTHDARADDGALLHMPQAPVALLDSPASLRAFLTLTKRLTFVLEETALLP
 TAERVARRLLMADGYDRLGSTRVRLVPEQLALLLSRTQVNLKDFEARGLL
 RLAYGEIELDFAGLRALARG"
 5934..6503
 /gene="RSP0806"
 /note="synonym: RS01901"
 5934..6503
 /gene="RSP0806"
 /function="miscellaneous; hypothetical/global homology"
 /note="Product confidence : probable
 Gene name confidence : hypothetical
 predicted by Codon_usage
 predicted by Homology
 predicted by Framed"
 /codon_start=1
 /evidence=not_experimental
 /transl_table=1
 /product="PROBABLE TRANSMEMBRANE PROTEIN"
 /protein_id="CAD17957.1"
 /db_xref="GI:17431278"
 /db_xref="GOA:O8XRM5"
 /db_xref="SPTREMBL:O8XRM5"
 /translation="MLRPHNPARTIMPLSDHLSYAAHYODGNITATHPFGPIVY
 AVAVLFRPVGLPGGAVATPALLILVAVTFRLDVFGLMGLIGMAVAVGH
 VAAHSTVAMLSVGLFVIGMIVQVGHVYEGRPKPAVDLAGLIVGFLRLAETVFA
 MARGALRDEAVASARARAAAPGKHAHA"
 complement(6511..6849)
 /gene="RSP0807"
 /note="synonym: RS01900"
 complement(6511..6849)
 /gene="RSP0807"
 /function="small molecule metabolism; energy transfer;
 electron transport"
 /note="Product confidence : probable
 Gene name confidence : hypothetical
 predicted by Codon_usage
 predicted by Homology
 predicted by Framed"
 /codon_start=1
 /evidence=not_experimental
 /transl_table=1
 /product="PROBABLE FERREDOXIN PROTEIN"
 /protein_id="CAD17958.1"
 /db_xref="GI:17431279"
 /db_xref="GOA:O8XRM4"
 /db_xref="SPTREMBL:O8XRM4"
 /translation="WPIVTVESCTQCKYTDCVAVCPMDCFHAGPNFLVDPPTCIDS
 ICAPCEPVGAIIHADVADQREFIALNAQLSRADWFLTVQVPLADHARAQYKD
 KRALIVTPA"
 complement(7004..8746)
 /gene="RSP0808"
 /note="synonym: RS01899"
 complement(7004..8746)
 /gene="RSP0808"
 /function="miscellaneous; hypothetical/partial homology"

Alignment Scores:
 Pred. No.: 5,17e-34 Length: 197050
 Score: 758.00 Matches: 182
 Percent Similarity: 60.98% Conservative: 68
 Best Local Similarity: 44.39% Mismatches: 136
 Query Match: 37.54% Indels: 24
 DB: 1 Gaps: 9

US-09-889-756a-2 (1-412) x AL646081 (1-197050)

Qy 12 AAlaAlaLeuAlaAlaAlaValAlaLeuValLeuSerSerCysGlyGlyGlyAspAla 31
 Db 24688 GCCGCGCTGCGCC-----GCCAGCTGCTGCGCCGCGGCAAG----- 24647
 Qy 32 AAlaGlnGlyGlnProAlaGlyArgGluAlaProAla-----ProValValGly 48
 Db 24646 -----CCGCGCGGTGGCGCCCGCCAGCGAGCGCGCGGTGCGC 24602
 Qy ValValThrValHisProGlnThrValAlaLeuThrValGluLeuProGlyArgLeuGlu 68
 Db 24601 GTGATGACCTGACGCGCGCGCGCGCTGATCCGATCCGAGCTGCGCGGCGACGCTC 24542
 Qy 69 SerLeuArgThrAlaAspValArgAlaGlnValGlyGlyIleIleGlnIlySarGluPhe 88
 Db 24541 CCTTCCTGCTCCCGATGTCGCGCCCAAGTCACAGCGCATCAAGCGCGCAATTC 24482
 Qy 89 GlnGlnGlySerIlyrValArgAlaGlnProLeuTyrGlnIleAspSerThrTyr 108
 Db 24481 CGGAGGCGACGACGGAAGCGGCGCGCTCTACAGATCGATCCGCGCACTAC 24422
 Qy 109 GluAlaAsnLeuGluSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaLys 128
 Db 24421 CAGCGCGCTCAGACAGACAGACGCGCGCTGCGCGCAAGGCCAGCGCACTCAAGACC 24362
 Qy 129 AlaAspAlaAspLeuAlaArgTyrLysProLeuValAlaAlaGluAlaValSerArgGln 148
 Db 24361 ACCGCGCTGAAGCGCGCGCGCACAGAGCTGTGCGATCCAGCGCGTCAAGCGGCG 24302
 Qy 149 GluTyrAspAlaAlaValThrAlaLysArgSerAlaGluAlaGlyValLysAlaAlaGln 168
 Db 24301 GACTAGACAGACCGCGCGCGCTGCGCGCACGAGGCCAGCGCGCGCGCGC 24242
 Qy 169 AlaAlaIleLysSerAlaGlyIleAsnLeuAsnArgSerArgIleThrAlaProIleSer 188
 Db 24241 GCCAAGTGAAGACGACGCGATCACTGCGCTACGCGCGGAGTGGAGTCCCATCTCC 24182
 Qy 189 GlyPheIleGlyGlnSerIlyrValSerGlnGlyThrLeuLeuAsnAlaGlyAspThrThr 208
 Db 24181 GCGCGATGGGACGATCCAGCGATGACCGCGCGCGCTGTCAACCGCAACGACCAACC 24122
 Qy 209 ValLeuAlaThrIleArgGlnThrAsnProMetTyrValAsnValThrGlnSerAlaSer 228
 Db 24121 TCCTGCGCACGATCCAGACGCTGCAACCCATCTAGTGAGCTGACCGCGCAGCGCC 24062
 Qy 229 GluValMetLysLeuArgArgGlnIleAlaGlnGlyLysLeuLeuAlaAlaAspGlyVal 248
 Db 24061 GCACCTGCGCTGCGCGCGCGCGATGGCGCGCGGATCTCCAGAAAGCGCGCCAC 24002
 Qy 249 IleAla---ValGlyIleLysPheAspAspGlyThrValTyrProGluLysGlyArgLeu 267
 Db 24001 GCGGCGACCGTGGCGCTGCTGTAAGAAGCGGACGCGATCCGCTCGAAGCGCAAGCTG 23942
 Qy 268 LeuPheAlaAspProValValAsnGlnSerThrGlyGlnIleThrLeuArgAlaAlaVal 287
 Db 23941 GAATTCTCGACGTGACGATGACGCAACACCGGCTCCGTGACGCTGCGCGGTGTTTC 23882

Qy 288 ProhnaAspGlnAsnIleuMetProGlyLeuTyValArgValIleuMetAspGlnVal 307
Db 23881 CCACACCCGAGCGCATCTGCTGCCGCGCATGTATGTGGCGCGCTGCTGCCGGAAGGC 23822
Qy 308 AlaValAspAsnAlaPheValValProGlnGlnAlaValThrArgGlyAla-----Lys 325
Db 23821 GTGAAGACGAGCGCCCTCTGCTGTCCGACAGCGCGCGCGGACAGACACCGGCAAG 23762
Qy 326 AspThrValMetIleValAsnAlaGlnGlyMetGluProArgGluValThrValAla 345
Db 23761 CCGTTCGCCCTACGTGTGTGGCAGACCGCAAGCTGCAGCCGCCGACCTGTGAACGAG 23702
Qy 346 GlnGlnGlnGlyThrAsnTrpIleValThrSerGlyLeuTyAspGlyAspLysValVal 365
Db 23701 CGCACCGGTGTGACAGGTGTGTGTGGCAGCGCGCTGCGCATGCGACCACTGTTG 23642
Qy 366 ValGlnGlyIleSerIleAlaGlyIleThrGlyAlaLys--LysValThrProLysGlu 384
Db 23641 GTGGAAGCGCTGCCCGTGC---GTGCCCGCGCGCGGTCAAGACACCGCG--- 23591
Qy 385 Trp-----AlaSerSerGlnAsnGlnAlaAlaProGlnSerGlyValGlnThr 401
Db 23590 TGGACGGCAGACCGCGACGTCGACCCCGCGCGCGCGCGCGCGCTGTGTGGACG 23531
Qy 402 AlaSerGlnAlaLysThrAlaSerGlnAla 411
Db 23530 GCCGTGCGGCGACCGCACCGCTGCCGCC 23501

RESULT 26
AL627267/c 240050 bp DNA linear BCT 06-JUN-2002
LOCUS Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
DEFINITION complete chromosome; segment 3/20.
ACCESSION AL627267 AL513382
VERSION AL627267.1 GI:16501740
KEYWORDS
SOURCE
ORGANISM
SALMONELLA enterica subsp. enterica serovar Typhi
SALMONELLA enterica subsp. enterica serovar Typhi
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
REFERENCE
AUTHORS
1 (bases 1 to 240050)
Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,
Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,
Sebaiha,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,
Conerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,
Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S.,
Jagels,K., Krogh,A., Larsen,T.S., Leachter,S., Moule,S., O'Goara,P.,
Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,
Stevens,K., Whitehead,S. and Barrett,I.B.G.
Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18
NATURE 413 (6858), 848-852 (2001)

TITLE
JOURNAL
MEDLINE
21534947
PUBMED
11677608
2 (bases 1 to 240050)
Parkhill,J.
Direct Submission
Submitted (25-OCT-2001) Submitted on behalf of the Salmonella
sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA, UK
E-mail: parkhill@sanger.ac.uk
COMMENT
Notes:
Details of S. typhi sequencing at the Sanger Centre are available
on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/S_typhi/).
Location/Qualifiers
1. 240050
/organism="Salmonella enterica subsp. enterica serovar
Typhi"
/mol_type="genomic DNA"
/strain="CT18"
/db_xref="taxon:90370"
213. 473
gene

CDS
/gene="STY0512"
/note="synonyms: rpmE2, ykGM"
213. 473
/gene="STY0512"
/note="Similar to Escherichia coli hypothetical 9.9 kDa
protein in int-*eah* intergenic region ykGM SW:YKGM_ECOLI
(P71302) (87 aa) fasta scores: E(): 1.6e-27, 74.4% id in
86 aa, and to *Listeria monocytogenes* 508 ribosomal protein
131 rpmE SW:RL31_LISMO (Q9ZH28) (81 aa) fasta scores: E():
7.9e-12, 48.1% id in 79 aa, and to *Borrelia burgdorferi*
508 ribosomal protein 131 rpmE or db0229 SW:RL31_BORBU
(O51247) (81 aa) fasta scores: E(): 7.8e-11, 43.6% id in
78 aa
Fasta hit to RL31_ECOLI (70 aa), 33% identity in 84 aa
overlap
Orthologue of E. coli YKGM_ECOLI. Fasta hit to YKGM_ECOLI
(87 aa), 74% identity in 86 aa overlap"
/codon_start=1
/transl_table=11
/product="putative 50s ribosomal protein L31 (second
copy)"
/protein_id="CAD04954.1"
/db_xref="GI:16501741"
/db_xref="SPREMBL:08XGM1"
/translation="MKPDHPVYRTVVFHDTSANEVYKVGSTIKTEREILDGVTPY
VTIDVSKSHPFYTGROKTFDSESSARFQKRFHFIKARG"
213. 452
/gene="STY0512"
/note="Plam match to entry PF01197 Ribosomal L31,
Ribosomal protein L31, score 48.40, E-value 1.6e-10"
479. 619
/gene="STY0513"
/note="synonym: rpmJ2"
479. 619
/gene="STY0513"
/note="Similar to *Gillardia cheta* chloroplast 50s
ribosomal protein 136 rpl36 SW:RX36_GURH (P28528; O46911)
(48 aa) fasta scores: E(): 1.6e-09, 56.5% id in 46 aa, and
to *Rickettsia prowazekii* 508 ribosomal protein 136 rpmJ
or rp456 SW:RL36_RICPR (Q9ZD87) (41 aa) fasta scores: E():
4.9e-08, 56.1% id in 41 aa"
/codon_start=1
/transl_table=11
/product="putative 50s ribosomal protein L36 (second
copy)"
/protein_id="CAD04955.1"
/db_xref="GI:16501742"
/db_xref="SPREMBL:08XJ38"
/translation="MOVLNSLRNAKORHPDQIVKRKGRLYVICTNPRFKAVGRKK
RR"
479. 601
/gene="STY0513"
/note="Plam match to entry PF00444 Ribosomal L36,
Ribosomal protein L36, score 23.10, E-value 2.2e-06"
complement (675. 1145)
/gene="STY0514"
complement (675. 1145)
/gene="STY0514"
/note="Orthologue of E. coli YLAC_ECOLI. Fasta hit to
YLAC_ECOLI (169 aa), 81% identity in 155 aa overlap.
Contains two potentially membrane spanning domains"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAD04956.1"
/db_xref="GI:16501743"
/db_xref="SPREMBL:08Z8U0"
/translation="MTEIORLSETIDDLNVREKRDNRPFSSIFKRGFLFIAMYA
AMFATLVMLOSETIGSVLIVLFLAFNGFPFDPAPAYHYNDIVDLRCVNGE
WYNTRVPTLITFILOSPOVDNEHKVQLQKMARKEGSLFYDIFTLARAEAR"
complement (1263. 1814)
/gene="STY0515"
/note="synonym: maA"

misc_feature
479. 601
/gene="STY0513"
/note="Plam match to entry PF00444 Ribosomal L36,
Ribosomal protein L36, score 23.10, E-value 2.2e-06"
complement (675. 1145)
/gene="STY0514"
complement (675. 1145)
/gene="STY0514"
/note="Orthologue of E. coli YLAC_ECOLI. Fasta hit to
YLAC_ECOLI (169 aa), 81% identity in 155 aa overlap.
Contains two potentially membrane spanning domains"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAD04956.1"
/db_xref="GI:16501743"
/db_xref="SPREMBL:08Z8U0"
/translation="MTEIORLSETIDDLNVREKRDNRPFSSIFKRGFLFIAMYA
AMFATLVMLOSETIGSVLIVLFLAFNGFPFDPAPAYHYNDIVDLRCVNGE
WYNTRVPTLITFILOSPOVDNEHKVQLQKMARKEGSLFYDIFTLARAEAR"
complement (1263. 1814)
/gene="STY0515"
/note="synonym: maA"

gene

```

CDS      complement(1263..1814)
        /gene="STY0515"
        /EC_number="2.3.1.79"
        /note="Similar to Escherichia coli maltose
        O-acetyltransferase maa SW:MAA_ECOLI (P77791) (182 aa)
        fasta scores: E(): 0, 79.7% id in 182 aa, and to Rhizobium
        meliolioti nodulation protein 1 modL SW:MODL_RHINE (P28266)
        (183 aa) fasta scores: E(): 4.6e-30, 47.5% id in 177 aa,
        and to Escherichia coli galactoside O-acetyltransferase
        SW:THGA_ECOLI () (203 aa) fasta scores: E(): 1.4e-28,
        40.8% id in 179 aa
        Fasta hit to THGA_ECOLI (203 aa), 41% identity in 179 aa
        overlap
        Orthologue of E. coli MAA_ECOLI; Fasta hit to MAA_ECOLI
        (182 aa), 80% identity in 182 aa overlap"
        /codon_start=1
        /transl_table=11
        /product="maltose O-acetyltransferase"
        /protein_id="CAD04957.1"
        /db_xref="GI:16501744"
        /translation="MSDEKQKIMAGALYCPDETLLCODRLAPOLIHOYVTTDEIN
        RKQALHGTGCEADAVIEPSPRCYGNIFLGHSFYANPCWALDVCPIHIGNCML
        AGVHIYATATHTPLDAVERNSGSEIGKEPTIIGNNVIIGRAVYVNGVITIGNVAVASRA
        VVTXVVPVVYVGNPARIKTL"
        /gene="STY0515"
        /note="PS00101 Hexapeptide-repeat containing-transferases
        signature"
        complement(1317..1403)
        /gene="STY0515"
        complement(1323..1376)
        /gene="STY0515"
        /note="Pfam match to entry PF00132 hexaped, Bacterial
        transferase hexapeptide (four repeats), score 16.40,
        E-value 0.68"
        complement(1377..1430)
        /gene="STY0515"
        /note="Pfam match to entry PF00132 hexaped, Bacterial
        transferase hexapeptide (four repeats), score 13.20,
        E-value 6.2"
        complement(1485..1538)
        /gene="STY0515"
        /note="Pfam match to entry PF00132 hexaped, Bacterial
        transferase hexapeptide (four repeats), score 17.30,
        E-value 0.37"
        complement(1593..2211)
        /gene="STY0516"
        /note="synonym: hha"
        complement(1993..2211)
        /gene="STY0516"
        /note="Fasta hit to YDGT_ECOLI (71 aa), 38% identity in 71
        aa overlap
        Orthologue of E. coli hha (HHA_ECOLI); Fasta hit to
        HHA_ECOLI (72 aa), 99% identity in 72 aa overlap"
        /codon_start=1
        /transl_table=11
        /product="haemolysin expression modulating protein"
        /protein_id="CAD04958.1"
        /db_xref="GI:16501745"
        /db_xref="SPTREMBL:Q9L654"
        /translation="WSDKPLTKTIDWLRLRCQITIDTLERYIEKNKTELSDNELAVFY
        SAADHRLAELTMNKKYDKIPSSVWKFIR"
        complement(2239..2613)
        /gene="STY0517"
        /note="synonym: ybaJ"
        complement(2239..2613)
        /gene="STY0517"
        /note="Orthologue of E. coli ybaJ (YBAJ_ECOLI); Fasta hit
        to YBAJ_ECOLI (124 aa), 95% identity in 124 aa overlap"
        /codon_start=1
        /transl_table=11
        /product="conserved hypothetical protein"
        /protein_id="CAD04959.1"
        /db_xref="GI:16501746"

```

gene	/db_xref="SPTRMBL:08XGK5"	
CDS	/translation="MDKSPSRPHRIADKLPCLTYHDDCLANFESNHWGNDPTSAV NLQNLBLEHTIATPALYKIKYINEDNKLIAOIDEYLDLDTFMLFSSYGINTODLQKMR SGNLFRCFVNATANPVSLSG" complement (3109..6258) /gene="STY0519" /note="synonym: acrb" complement (3109..6258) /gene="STY0519" /note="fasta hit to ACRP_ECOLI (1034 aa), 78% identity in 1033 aa overlap"	
Alignment Scores:		
Pred. No.:	8.81e-34	Length: 240050
Score:	756.00	Matches: 166
Percent Similarity:	60.70%	Conservative: 78
Best local Similarity:	41.29%	Mismatches: 128
Query Match:	37.44%	Indels: 30
DB:	1	Gaps: 5
US-09-889-756A-2 (1-412) x AL627267 (1-240050)		
QY	11	AlaAlaAlaLeuAlaAlaAlaValAlaLeuValLeuSerSerCysGlyIlyGlyValAsp 30
DB	7444	GCGGTCGTTCTGATGCTCTCAGGCACTTACCGCTTAACAGATGTGACGACAAACAGAC 7385
QY	31	AlaAlaGngIyGlyGlnProAlaGlyArgGluAlaProAlaProValIyValIyVal 50
DB	7384	---CAGCAAGCGGCGCAG-----CAGATGCCAGAAAGTTGGGGTTGTC 7346
QY	51	ThrValHisProGlnThrValAlaLeuThrValGluLeuProGlyArgLeuGluSerLeu 70
DB	7345	ACACTAAACCGGAACACATCGCAATATCAACATGACTCCGGGTCTTACCGTTTAC 7286
QY	71	ArgThrIlaAspValaArgAlaGlnValGlyIleIleGlnIlyAsArgLeuPheGlnGlu 90
DB	7285	CGIATCGCGGAAGTTCCCGCCGACAGTAAAGCGGCACTTATCTGAAAGCGTAAATTTCGTGAG 7226
QY	91	GlySerTyValaArgAlaGlyGlnProLeuTyGlnIleAspSerSerThrTyGluAla 110
DB	7225	GGAATGATATCGAAGCGGAGTCTCTCTATCAGATTATCTCGGACTTACAGGCG 7166
QY	111	AsnLeuGluSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaIlyAsAlaAsp 130
DB	7165	ACTTACGACAGCGCTAAGCGCATCTGGCAAAAGCGCAGCGCCGCGAATATGCGTGA 7106
QY	131	AlaAspLeuAlaArgTyIlyIysProLeuValAlaAlaGluAlaValSerArgGlnGluTy 150
DB	7105	CTGACGGTGAAGCGCTTATCAAAAGCTCTGGGTACGACATCATCAATGACGAGAAATAC 7046
QY	151	AspAlaAlaValThrAlaIlyAsArgSerAlaGluAlaGlyValIlyAsAlaAlaGlnAlaAla 170
DB	7045	GATAGAGCGCTGGCTGACGCGCAACAGCACTGCGCGCTTGTCCGACGAAAGCCGCC 6986
QY	171	IleIysSerAlaGlyIleAsnLeuAsArgSerArgIleThrAlaProIleSerGlyPhe 190
DB	6985	GTTGAAACCCGACGATCAACTGCGGCGTATACCAAAAGTCACTCACGATTAAGCGGTCT 6926
QY	191	IleGlyGlnSerIlyValSerGluGlyIlyThrLeuLeuAsnAlaGlyAspThrThrValLeu 210
DB	6925	ATTGGTAAAGTCGTCCTGTAACGGAAGGCGCCTGGTACGAACGCTCAGCGCGCTG 6866
QY	211	AlaThrIleArgGlnThrAsnProMetTyValAsnValThrGlnSerAlaSerGluVal 230
DB	6865	GCGACATGACGACGCTGAGACCTATTATGATGATGACCAAGTCACATGACATGATTC 6806
QY	231	MetIlyLeuArgArgGlnIleAlaGluGlyIlyLeuLeuAlaAlaAspGlyValIleAla 250
DB	6805	CTGGCGCTGAAGCAGCAGAGCTGGCAATAGTTCCTGTAACGAGAAACGCAAAAGCAAG 6746
QY	251	ValGlyIlyLeuSphaAspGlyThrValTyProGluIlyGlyArgLeuLeuPheAla 270
DB	6745	GTCGATCTGCTGACAGCGCGGATCAATTCCTCCGACGTCGCTTGAATTCCTCC 6686

```

Qy 271 AspprovalValaAsnGluSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnasp 290
    |||.....|
Db 6685 GACGTGACCGCTTGACCAAGACCGGGCTCTTACTTTGGCGGCATCTTCCCTAACCG 6626
    |||.....|
Qy 291 GlnAsnIleLeuMetProGlyLeuTyValArgValleuMetAspGlnValAlaValAsp 310
    |||.....|
Db 6625 GATCACACCTTATGTCGAGAAATGTCCTGCGCGACGATCGACGAGGACCAAAACCG 6566
    |||.....|
Qy 311 AsnAlaPheValValProGlnGlnAlaValThrArgGlyAlaValAsp-----ThrVal 328
    |||.....|
Db 6565 ACCGATTAATCTGCTTCCACCAACAGGCGCTTACCGCTACTCCACCGGCGATGCCACGCTG 6506
    |||.....|
Qy 329 MetIleValAsnAlaGlnGlyGlyMetGluProArgGluValThrAlaAlaGlnGln 348
    |||.....|
Db 6505 CTGGGTGGTGGCGCTGATACCAAAAGTGAACCCGCCAATCGTCGCAAGCAGCGCATC 6446
    |||.....|
Qy 349 GlyThrAsnTrpIleValThrSerGlyLeuLysAspGlyAspLysValValValGlnGly 368
    |||.....|
Db 6445 GGCGATAGTGGCTGCTGATGACGGGTTGAAGCGGCGACCGCGTATGTCGTCAGCGGG 6386
    |||.....|
Qy 369 IleSerIleAlaGlyIleThrGlyAlaLysValThrPro----- 382
    |||.....|
Db 6385 CTG-----CAAAAGTACGTCTCTGCGCGACAGGTTAAAGTG 6350
    |||.....|
Qy 383 -----LysGluTrpAlaSerSerGluAsnGlnAlaAlaAlaPro 395
    |||.....|
Db 6349 CAGGAATTACCGCGGATACCAACAGCAACCGGATGATCAACCTGCTCAGGCC 6250
    |||.....|
Qy 396 GlnSer 397
    |||.....|
Db 6289 AGTCT 6284

RESULT 27
AB016842 300029 bp DNA linear BCT 21-MAR-2003
LOCUS Salmonella enterica subsp. enterica serovar Typhi Ty2, section 9 of
DEFINITION 16 of the complete genome.
ACCESSION AB016842 AB014613
VERSION AB016842.1 GI:29138362
KEYWORDS
SOURCE
ORGANISM
Salmonella enterica subsp. enterica serovar Typhi Ty2
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
1 (bases 1 to 300029)
Deng, W., Lion, S.-R., Plunkett III, G., Mayhew, G. F., Rose, D. J.,
Burland, V., Kodoyianni, V., Schwartz, D. C. and Blattner, F. R.
Comparative Genomics of Salmonella enterica Serovar Typhi Strains
Ty2 and CT18
J. Bacteriol. 185 (7), 2330-2337 (2003)
JOURNAL MEDLINE 225131367
PUBMED 12644504
2 (bases 1 to 300029)
Deng, W., Lion, S.-R., Plunkett, G. III, Mayhew, G. F., Rose, D. J.,
Burland, V., Kodoyianni, V., Schwartz, D. C. and Blattner, F. R.
Direct Submission
Submitted (25-SEP-2002) Laboratory of Genetics, University of
Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
source
1. 300029
location/Qualifiers
/organism="Salmonella enterica subsp. enterica serovar
Typhi Ty2"
/mol_type="genomic DNA"
/strain="Ty2"
/serovar="Typhi"
/sub_species="enterica"
/db_xref="taxon:209261"
complement(170..1524)
/gene="ybhW"
/locus_tag="t2339"
/note="corresponds to STY0569 from Accession AL513382:
Salmonella typhi CT18"
gene
CDS
/pseudo
complement(1620..2867)
/locus_tag="t2340"
complement(1620..2867)
/locus_tag="t2340"
/note="corresponds to STY0568 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="probable metabolite transport protein"
/protein_id="AA069933.1"
/db_xref="GI:29138363"
/translacion="MSETKSGIDYWKQIVVNSLGVWIMIVRTVLPYPIQASL
DANSNAEIGIAFFEFAYCSMOIPGILVDKFGOKIMAGFTLPIGTLIAKANG
LAMIYSGIWAAGGCAFFESAYSLSANVPORALANAINSGSAGIGGIGSS
VIKNSMAMQNTLYVAALIVMLCYFTLIERGDSADAEKOTQYTPDEKAP
LFGSLCSVFILFTCTGYLYLVTPSLQTERGDSATGLASALVAVVGPAL
FESHLSKRFNSRVKYLIGLEIVAAAMLATVLSPTTLMVSLTLGLGLKADP
LISFVSQASAKSLGRASLSENFPMSSSAVAVPTLTGIFSDVGSKEISFVISA
TGTLIFAVTLVKKKATORIASA"
complement(2933..3811)
/gene="garR"
/locus_tag="t2341"
complement(2933..3811)
/gene="garR"
/locus_tag="t2341"
complement(2933..3811)
/locus_tag="t2341"
/note="corresponds to STY0567 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="2-hydroxy-3-oxopropionate reductase"
/protein_id="AA069934.1"
/db_xref="GI:29138364"
/translacion="MKLGFGLGIMSPMIALNAPAGHOLHTTIGVPADELISLAV
NETHARQVTEFPADIFIMVPTPQVEVLEGECAKTSIQGTIDMSISPIGTR
PAQVWEMDADYLDAPVSGEIGAREGTSIMWGSQVDRKPLFDILGKNTLVG
GNDGQTCVKANQIYVALNIEAVSALVFASKAGADPVRVROALMGFPASSRILEVG
ERMNRTPFEGFKIALHOKDLNLALQSAKALNLNPTATCOELFTNCANGSOLDH
SAMVOLBELMANHKLS"
complement(3912..4688)
/gene="hyi"
/locus_tag="t2342"
complement(3912..4688)
/gene="hyi"
/locus_tag="t2342"
/note="Tccorresponds to STY0566 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="hydroxypyruvate isomerase"
/protein_id="AA069935.1"
/db_xref="GI:29138365"
/translacion="MLRFSANLSMLTFEYPLERFPDKAASGRGVFEMFPYDDIV
LKKRLRNNLEHTLHNPAGDMVAGEGACIGRESEFPDGAARIRVAPALGNKI
NLCVGGKPSGSAFTEIYDVLVENIRVAAANMLAEDILLLLEPINFDPMPHLGTQO
ALVLIKIDGSDNFKIYDIYHOMRMEGELTOTWTWADKIGHQIADNPRRGEGTGE
INVDPIFNVIKOSDYOGWCCEYKPLTTABAGSLINQYV"
complement(4701..6482)
/gene="gcl"
/locus_tag="t2343"
complement(4701..6482)
/gene="gcl"
/locus_tag="t2343"
complement(4701..6482)
/locus_tag="t2343"
/note="Tccorresponds to STY0565 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="glyoxylate carboxylase"
/protein_id="AA069936.1"
/db_xref="GI:29138366"
/translacion="WAKRRAVDAMVYLEEGITTAFGVGAALNPFYSAMRKGGIR
HILARHVEGASHMAGETRAAGNIGCLCTSGPAGDMLTALAYSASADSIPLICIG
QAPRARLHKEDFOAVDIEALAKVSKMAVTVREAAVPRVLQQAFLHMRSGRGPVLV

```



```

Db      51367 CTGGCCCTGAAGCAGAGCTGCGCAAAATGTTGCTGAAGAGCAAAACGCGCAAGCGAAG 51426
Qy      251 ValGlyIleLysPheAspAspGlyThrValTyrProGlyIuLysGlyLeuLeuPheAla 270
Db      51427 GTCATCTCGTGACAGAGGAGGTATCAAAATTCCTCCGACGTCGCGTACCGCTTGAATTCTCC 51486
Qy      271 AspProValValaAngLysSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAsp 290
Db      51487 GACGTGACCGTTGACCAAGACCGGGGCTCTATTACTTGGCCGCGCATCTTCCCTAACCCG 51546
Qy      291 GlnAsnIleLeuMetProGlyLeuTyrValArgValLeuMetAspGlnValAlaValAsp 310
Db      51547 GATCACACCTTATGTCAGGAGATGTTCTGTCGCGACGTCGACGAGGAGGACAAACCG 51606
Qy      311 AsnAlaPheValValProGlnAlaValaThrArgGlyAlaLysAsp-----ThrVal 328
Db      51607 ACGGCATTACTGTTCCCAACAGGCGCTTACCGCTACTCCACGCGCGATGCCAGGTG 51666
Qy      329 MetIleValaAsnAlaGlnGlyGlyMetGluProArgGluValaThrValaIaGlnGln 348
Db      51667 CTGGGTGGCGCTGATTAACAAAGTGAAACCGCCCAATCGTCGCAAGCCAGGCGATC 51726
Qy      349 GlyThrAsnThrPleValaThrSerGlyLeuLysAspGlyAspLysValaValaGluGly 368
Db      51727 GGCATTAAGTGGCTGCTGACTGACGAGGTTGAAGCGGCGACCGCGTACTGTCAGCGGG 51786
Qy      369 IleSerIleAlaGlyIleThrGlyAlaLysLysValaThrPro----- 382
Db      51787 CTG-----CAAAAAGTACGCTCTGCGCGCACAGCTTAAAGTG 51822
Qy      383 -----LysGluThrPalaSerSerGluAsnGlnAlaAlaPro 395
Db      51823 CAGAAATTACCGCGGATTAACAAACAGCAGCAGCGTGATCAACCTGTCAGGCC 51882
Qy      396 GlnSer 397
Db      51883 AGGTCT 51888

RESULT 28
PSENVCD      5600 bp      DNA      linear      BCT 16-AUG-1995
LOCUS      PSENVCD      5600 bp      DNA      linear      BCT 16-AUG-1995
DEFINITION      Pseudomonas aeruginosa mexa and mexb genes, complete cds and outer
                  membrane protein (oprM) gene, partial cds.
VERSION      L11616
KEYWORDS      L11616.1 GI:438852
SOURCE      homologe.
ORGANISM      Pseudomonas aeruginosa
                  Pseudomonas aeruginosa
                  Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
                  Pseudomonadaceae; Pseudomonas.
REFERENCE      1 (bases 1 to 5600)
AUTHORS      Poole, K., Heinrichs, D.R. and Neshat, S.
TITLE      Cloning and sequence analysis of an EnvCD homologue in Pseudomonas
                  aeruginosa: regulation by iron and possible involvement in the
                  secretion of the siderophore pyoverdine
JOURNAL      Mol. Microbiol. 10 (3), 529-544 (1993)
MEDLINE      95058196
PUBMED      7968531
REFERENCE      2 (bases 1 to 5600)
AUTHORS      Poole, K.
TITLE      Direct Submission
JOURNAL      Submitted (02-JAN-1994) Keith Poole, Microbiology and Immunology,
                  Queen's University, Kingston, Ontario, Canada
COMMENT      Original source text: Pseudomonas aeruginosa (strain CD10) DNA.
FEATURES
    source
        1. .5600
            /organism="Pseudomonas aeruginosa"
            /mol_type="genomic DNA"
            /strain="CD10"
            /db_xref="taxon:287"
            403. .1554
            /gene="mexa"
            403. .1554
    CDS

```

```

/gene="mexa"
/function="possible function in pyoverdine secretion"
/note="envc homologue; ORF a"
/codon_start=1
/evidence=experimental
/transl_table=11
/protein_id="AA474436.1"
/db_xref="GI:438853"
/translation="MORTPAMRVLVPAALVAISALSGCCSKSEAPPAQTPPEVGLITLE
AQTIVNTLPGRTNAPRIAEVRPQVNGIILKLFKESGSDVKAGQOOLYQIDPATYEA
VGSANGLASTQSOAORXYKLIVADQSVKQVADANAAYQSAVAEQARINLRITYV
LSPTSGRIGHSATBGLVYNGQANMATVQOQDPIYVDTQSTALIRRLRLASQ
LERAGDAAAVSILKLEDGSOYPLEGRLESEVSDGTSVTLRAEPNNBELPQM
FVAHQLOEQKQKAILAPQOGVTRDLKGAATLVAANQNEVLRLVKADRVIDKMLV
TEGLNAGDKIITTEGLQFVQPGVEVKTVPAKNVASAKADAPAKTDSKG"
1570. .4710
/gene="mexb"
1570. .4710
/gene="mexb"
1570. .4710
/function="possible role in pyoverdine secretion"
/note="envd homologue; ORF b"
/codon_start=1
/evidence=experimental
/transl_table=11
/protein_id="AA474437.1"
/db_xref="GI:438854"
/translation="MSKFFIDRPIFAMVILVIMLAGLSILSPVNOYPAIPAPA
VOVSPGASAEVTDYVVOVYBOONGIDRLRISSESSNGMTTTPVEQGDPI
AQOVONKLOLAPLPLPOBYROGIVYTRAKVNFVWNGVSDGSMTEKEDLSNYIS
NIQDPVSRKRGVDFOVPSQYSMRIMLDPKAKNSVQLTFGDSALIQONVOISQO
LGGLPAKGGQOLNATIIIGKRLQTAQFENILKTVPDGQSVLAKVADVLGQDYS
INAQFNSPAGSIAKILATCANALDPAKALROTIANLLEPMPQMKRVVYDTPVVS
ASJHEVKTGEBAILFLVIMVFLQNFATLPIPLTAPVPLVLTGTEGLAAFEFSINT
LIMGEVATIGLVDPAIVVENVERVMAEGSLPBEAPKASMOGOLGALVAMVLS
AVLIPMAFRGSGTGYVROSTITVSAMLSVVALIIPALCATMILKPIEKDGHGB
KGFPRGNMFLSTHGTGRGVASILKRAPIPLIYIVVAGIMKFTRIPIRAPLBD
EDGCVLFAOVQTPGSSAERTQVVDMSREPLILEKSSSVSVFTVGFNPAAGQSS
GMAFIMLKPEWERPGENSVFELAKQOMFSPKQAMVFAFAPSYLEGNATGFL
FLDOAGVGHVILLQARKFLMLAQNPAQYRPMGMSDEPOYKLEIDDEKASALGV
SLADINVSIVIAMSSVYNDFIDRGVRKRYLGGPRPARNPDLISKVYVBNKGEMV
PNNAPATGKREKVSPTKLEBYNVPAMEIIGEPAPIGLSSGMAVAEETIYQIPGVGV
SWTGLSEERLSQAPALYALSLVFLTALALYESMSIPFSVMLVPIPGVALLA
TSNRGLSNDVFPQVGLTTTIGLAKNALIVEBAKHEGKGIVEAALBACNRRLP
IWNVSLAFLIGVPLAISTGAGSGOHALIGTGVIGKMTATVLAIFWVPLFYAVVSTL
FXDBASKQASVEKQG"
4712. .5600
/gene="oprM"
4712. .5600
/gene="oprM"
/note="ORF; formerly called oprX"
/codon_start=1
/codon_start=1
/transl_table=11
/product="outer membrane protein"
/protein_id="AA474438.1"
/db_xref="GI:833173"
/translation="MKRSFLSLVAAVVLSGCSLIPDYQRPAPVAAYPOGAYGNG
TGAAYPAADIGWBEFPRPOLQOLIGVALENNRDLRVALNTEFAAYRIRQADLF
PRIGVSSGTRRPLPDLSTTSPALISSQVGLTGTVALDLRGRSLRDALDEY
LATBOARSAQTLIVASVATYATITUKADQOLQDTDTCTYOKSPULTRSDYVGA
SAULDRQAQTVAGARATTAQYTRLVAAQDONALVLLSGSPANLPGLDLDOTLLE
VPAQLPDLIDRRPDLIEAE"
BASE COUNT      1004 a      1845 c      1766 g      955 t
ORIGIN
ALIGNMENT Scores:
Pred. No.:      9.79e-36      Length:      5600
Score:      753.00      Matches:      174
Percent Similarity:      58.60%      Conservative:      68
Best local Similarity:      42.13%      Mismatches:      125
Query Match:      37.30%      Indels:      46
DB:      1      Gaps:      10
US-09-889-756a-2 (1-412) x PSENVCD (1-5600)

```

```

Qy      8 AlaMetArgAlaAlaAlaLeuAlaAlaValAlaLeu---ValLeuSerSerCysGly 26
Db      418 GCCATGGGTATACGTTCCGGCCCTGCTGTCGGATTCGGCCCTTTCGGGGTGGCGA 477
Qy      27 LysGlyValThrProLysGlyGlnProAlaGlyArgGluAlaProAlaProVal 46
Db      478 AAAAGC-----GAGGCGCCCGCCCGCGC 501
Qy      47 -----ValGlyValThrValHisProGlnThrValAlaLeuThrValGlu 62
Db      502 CAAAGCGCGAGGTCGGATCTGACCTGGAGCGGACGAGCGTGAACCTGAATACCGAG 561
Qy      63 LeuProGlyArgLeuGluSerLeuArgThrAlaAspValAlaArgAlaGlnValGlyIle 82
Db      562 CTGCGGGGCGGACCAATGCTTCGCGATCCGCGAGTGGCTCCCGCAGTGAAGCGATC 621
Qy      83 IleGlnLysArgLeuPheGlnGlySerTyValAlaArgAlaGlyIleProLeuTyGln 102
Db      622 ATCTCAAGCGCTCTCAAGAGGACGACCTCAAGCGCGGCGAGCAGCTTACACG 681
Qy      103 IleAspSerSerThrTyGluAlaAsnLeuGluSerAlaArgAlaGlnLeuAlaThrAla 122
Db      682 ATCAACCCCGCCACTCAAGAGCGGACCTACCAAGCGCGCAACCTGGCTTCAGC 741
Qy      123 GlnAlaThrLeuAlaLysAlaAspAlaAspLeuAlaArgTyIleProLeuValAla 142
Db      742 CAG-----GAAAGCGCGGCTCAAGCGCTGCTGTCGCGAC 780
Qy      143 GluAlaValSerArgGlnGlyTyArgAlaAlaValThrAlaLysArgSerAlaGluAla 162
Db      781 CAGCGCTGTGACCAAGCAGCAGTAC-----GCCGACGCCCAAGGCC 819
Qy      163 GlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsnArgSerArg 182
Db      820 GCTTACTGCACTCAAGCGCGGCTGTGAGCGGCGGATCAACCTGCGCTACACCAAG 879
Qy      183 IleThrAlaProLysSerGlyPheIleGlyIleSerLysAlaSerGlnGlyThrLeuLeu 202
Db      880 GTGCTGTGCGGATCTCCGCGGATGCGCGCTTCGCGGAGCGGACCAAGCGCGCTGTG 939
Qy      203 AsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAsnProMetTyValAsn 222
Db      940 ACCAAGCGCGCAAGCGGACGCGGATGTCGACGAGCTGACCGATCTAGCTGAC 999
Qy      223 ValThrGlnSerAlaSerGluValMetLysLeuArgArgGlnIleAlaGluGlyLysLeu 242
Db      1000 GTCAACCGCGCTCCACCGCGCTGCTGCGCGCGGCAACGCGGCGCGCGCTGTG 1059
Qy      243 ---LeuAlaAlaAspGlyValIleAlaValGlyIleLysPheAspGlyThrValTy 261
Db      1060 GAGCGCGCGCGGCAACCGCGGCAAGGTCTCTCGAAGCTGAGGAGCGGTAGCAATAC 1119
Qy      262 ProGlnLysGlyArgLeuLeuPheAlaAspProValValAsnGluSerThrGlyIle 281
Db      1120 CCGGTGAAAGTGGCTCTGCAATCTCGAGGTTCTCGTCAAGGACGACCGGCTCGT 1179
Qy      282 ThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetProGlyLeuTyValArg 301
Db      1180 ACCATTCGCGCGCTGTTCCCAACCGCAACGAGTGTGCGCGGCAATGTTGTGCAC 1239
Qy      302 ValLeuMetAspGlnValAlaValAspAsnAlaPheAlaValProGlnGlnAlaValThr 321
Db      1240 GCGGAGTTCGAGGAAGCGCTCAAGACGACCTCTCGTCCGCGAGGAGGCTGAC 1299
Qy      322 ArgGlyAlaLysAsp-----ThrValMetIleValAsnAlaGlnGlyMetGluPro 339
Db      1300 CGGCACTCAAGGCGCGGCTGCTGTGTAACCGCGCGGCAAGAGTGCAGCTG 1359
Qy      340 ArgGluValThrValAlaGlnGlnGlnGlyThrAsnTrpIleValIleThrSerGlyLeuLys 359
Db      1360 CGGGTATCAAGGCGCGGCTGATCGCGCAAGAGTGTGTACCGAAGGCGCTGAC 1419

```

```

Qy      360 AspGlyAspLysValValValGluGlyIleSerIleAla-----GlyIleThrGlyAla 377
Db      1420 GCCGCGCAACAATCATTACCAAGGCGCTGTCGACGCGGCTGTGAG---GTG 1476
Qy      378 LysLysValThrProLysGluTrpAlaSerSerGluAsnGlnAlaAlaAlaProGlnSer 397
Db      1477 AAGACCGTGGCGGGAAGATGTGCTCGCGGAGGAAGCGACGCGCTCCG----- 1530
Qy      398 GlyValGlnThrAlaSerGluAlaLysThrAlaSerGlu 410
Db      1531 -----GCCAATCCGACGACGAG 1548

RESULT 29
AE004479      13987 bp      DNA      linear      BCT 19-FEB-2003
LOCUS
DEFINITION . Pseudomonas aeruginosa PA01, section 40 of 529 of the complete
genome.
ACCESSION
AE004479 AE004091
AE004479.1 GI:9946272
VERSION
KEYWORDS
SOURCE
ORGANISM
Pseudomonas aeruginosa PA01
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE
1 (bases 1 to 13987)
Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,
Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,
Westbrock-Madman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
Polger,K.R., Kas,A., Laidig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Sater,M.H.,
Hancock,R.E.W., Lory,S. and Olson,M.V.
Complete genome sequence of Pseudomonas aeruginosa PA01, an
opportunistic pathogen
Nature 406 (6799), 959-964 (2000)
JOURNAL
MEDLINE
20437337
10984043
2 (bases 1 to 13987)
Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,
Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,
Westbrock-Madman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
Polger,K.R., Kas,A., Laidig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Sater,M.H.,
Hancock,R.E.W., Lory,S. and Olson,M.V.
Direct Submission
Submitted (16-MAY-2000) Department of Medicine and Genetics,
University of Washington Genome Center, University Of Washington,
Box 352145, Seattle, WA 98195, USA
3 (bases 1 to 13987)
Pseudomonas aeruginosa Community Annotation Project. (PseudocAP)
Direct Submission
Submitted (04-FEB-2003) Department of Molecular Biology and
Biochemistry, Simon Fraser University, 8888 University Dr.,
Burnaby, British Columbia V5A 1S6, Canada
COMMENT
This represents the February 3, 2003 version of the continually
updated, reviewed, Pseudomonas aeruginosa PA01 genome annotation,
from PseudocAP (see http://www.pseudomonas.com for latest updates
and links to alternate annotations). PseudocAP is coordinated by
Fiona S.L. Brinkman (Simon Fraser University, Canada) and Robert
E.W. Hancock (University of British Columbia, Canada). We welcome
submission through www.pseudomonas.com of any proposed changes.

'Protein name confidence' is used to rate our confidence of the
accuracy of the protein name.
Class 1: Function experimentally demonstrated in P. aeruginosa.
Class 2: Function of highly similar gene experimentally
demonstrated in another organism (and gene context consistent
in terms of pathways its involved in, if known).
Class 3: Function proposed based on presence of conserved amino
acid motif, structural feature or limited sequence similarity
to an experimentally studied gene.

```

Class 4: Homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences.

FEATURES	Location/Qualifiers
source	1..13987 /organism="Pseudomonas aeruginosa PAO1" /mol_type="genomic DNA" /strain="PAO1" /db_xref="taxon:208964"
gene	69..680 /locus_tag="PA0417" /note="synonym: chpE"
CDS	69..680 /locus_tag="PA0417" /note="Protein name confidence: Class 3 (function proposed based on presence of conserved amino acid motif, structural feature or limited sequence similarity to an experimentally studied gene)" /codon_start=1 /transl_table=1 /product="probable chemotaxis protein" /protein_id="AA03806.1" /db_xref="GI:9946273"
gene	/translation="MLAIFLALLFGFAFNVSGAVSETLRGLTGGRPALIVQGS SLIDAVWALLGLTGLALLGVQVRIPLTACAAVYLAWLGVQGLRDVMSPLAEDA GEOGRNAGAGAAISLSNPKVYWGALAGIVDGTQAOGLVFPAGFMSLI WCCCALVMDLRNTSLFMHRSVYAGGVLLGLAGALRGL"
CDS	complement(688..2103) /locus_tag="PA0418" complement(688..2103) /locus_tag="PA0418" /note="Protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences)" /codon_start=1 /transl_table=1 /product="hypothetical protein" /protein_id="AA03807.1" /db_xref="GI:9946274"
gene	/translation="WNTYAQPTSQRIAMDPVYFDKRIPLVDVGLVIRYAGRFPO LASALETLERLDDPEDFPALGOLESLSGLALVERLDEDSLNIYVGAHPSOR LIDALAAQGHNAARTLROENAEAGAAALPAAPAGLPAASAYICLEHAHAPLALIA RLPESDSRLDLRLQWPRQLPDGAETQCALCRASADGLHAWMTSGSPYVW LHRSRDLASTPELPLPPLPAQSLRQRTPELAIGLADGLLADRLFLYFPGQ GNDBPLFEVYQARSISNPVAVFTTPDQVCLISRGQFEMREARLQPLPFVGD LPADAPQTSAPGRTIVLIREPTCEGLDSCIRSHSLDGLAGPYLKIQAALGML LLAWLEPHRSRDLAQLMHESGRALRIRYGLDLBGLQHMALPDGRIVVGHANH VDLGTFSLRQLRRRLAFA"
CDS	complement(2138..2860) /locus_tag="PA0419" /note="synonym: Y99J" complement(2138..2860) /locus_tag="PA0419" /note="Protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences)" /codon_start=1 /transl_table=1 /product="conserved hypothetical protein" /protein_id="AA03808.1" /db_xref="GI:9946275"
gene	/translation="WRLSRFFVDDPLSGOHDLPBAOAHYIGRVLRHATPAVOLFDG SGREYIGELVGVKKNVLEBOPAGQPSRLIHGCGISGRBMDMAIOKTEIG VAETTPYSECEVRLKDERADKRLQHRQVAISACECCGSSPLIHPPLADWLR GRDDLRLVHVAAPLASHAPPSRLAVLVEGGLSDAEVEQAQAAGFHSARLGPV LRTEAPVVALVAQOLMGDA"
CDS	complement(2953..4356) /gene="b10a" /locus_tag="PA0420" complement(2953..4356) /locus_tag="PA0420" /note="Protein name confidence: Class 2 (high similarity

gene	to functionally studied protein); Alt protein name: 7,8-diamino-pelargonic acid aminotransferase DAPA aminotransferase"
CDS	/codon_start=1 /transl_table=11 /product="adenosylmethionine-8-amino-7-oxononanoate aminotransferase" /protein_id="AA03809.1" /db_xref="GI:9946276"
gene	/translation="MGLNADMQRDLNVLHMPCTOKMDHERLPIPIRGRGWLDEF EGKRYIDAVSSWVWVPGHANPRINIKQVNDOLSHVILAGSFHGVETLSRIYVI TPGLDRVRYADSGSAGIEVALKMSYHFLNLSRPRPKRVITLTNSHGTITAAVSG DVALFTETYSLLDITKVPSPDCELRPDGMCWESRRNFAMHRTLAGHDEILAV IVEPLIQAGCMRMHYPVYLKLREACDRYVHLIDELAVGRTGTMPACBOAGIA PDFLCLSKALTGCVLPMSAVLTSETTYRCGYDDYOTLRAFISHYTGNIACAAALA TLDFIEEDRYEAMRALSTMARATNLADHPHVAEYVOTGVLAIEVMODKASRTY PMOERGLKVFOHGLRGALRLRGLSVVYLPVYITPEQIDFLAEVASSGIDIAITRD AVSVANSDFRPHDRDPE"
CDS	4530..6020 /locus_tag="PA0421" 4530..6020 /locus_tag="PA0421" /note="Protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences)" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AA03810.1" /db_xref="GI:9946277"
gene	/translation="MSLRMLRAAFLMLGVFSAVALGKOKPTAIVWAGLAGLSNAY ELGKDQVIVLELARPQVGGSLATSEWGNQKQVPTLNAVYDTEFLKVPAPDFVR TMYLIDGLYSSSDILAKOPNVAAALDKRESTLDDLSAISPLNPASSNTLPALDO MMARWLDKNTLSPTARLILNORIRSGYRPSLSILYLAQOARVAGVDRDLRAR LPGSGVLAABVYKQIKITIKTSKVSISYQAKGVAVKASSETYKADYVYLAIPLAL GQIQMTPSLSGTQMSALKGTNYGWRDQILKFKRPVWDDKSRISGETFSQDGLMTV EPALIKGANLINLSGDNARVILQAFDRQWVDVLLRMKFFYPMKGAFAGYAIRYS ADPGTGSYLAQPGQVTFEWRIMEOPLSRVAGCHTDALVYGTTEGALRSKRAAS QVRDLYAGTKPTVEGAPVMWAKAEKVAFTTGGSKNPFAMFTLFD"
CDS	6201..6770 /locus_tag="PA0422" 6201..6770 /locus_tag="PA0422" /note="Protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences)" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AA03811.1" /db_xref="GI:9946278"
gene	/translation="MOWRNTSRGYVFSFLFHWGSAVVGFLGFLGMBELSYDDPW YHPAPALHKSIGILALIALVRYVWRVSPPPAPANHGGLTVASGLALYGLL AVIVAGTLISTADGEPISYGWFSPVTTSLGFLPDQADAGEIHLIYLAALVPAVLHA LAALKHFFVDRDPTLKRMLGRSSNDQPR"
CDS	6782..7357 /locus_tag="PA0423" /note="synonym: yceI" 6782..7357 /locus_tag="PA0423" /note="Protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences)" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AA03812.1" /db_xref="GI:9946279"

Alignment Scores:
Pred. No.:
Score:
Percent Similarity:

3.24e-35 Length: 13987
753.00 Matches: 174
58.60% Conservative: 68

gene /mol_type="genomic DNA"
/strain="LT2; SGSC 1412; ATCC 700720"
/db_xref="ATCC:700720"
/db_xref="taxon:99287"
RBS /note="LT2"
70..900
/gene="cof"
/note="synonym: STM0457"
70..75
/gene="cof"
/note="putative RBS for cof; RegulonDB:STMS1H000802"
82..900
/gene="cof"
/note="similar to E. coli orf, hypothetical protein (AAC73549.1); Blastp hit to AAC73549.1 (276 aa), 81% identity in aa 5 - 276"
/codon_start=1
/transl_table=1
/product="putative hydrolase"
/protein_id="AAL19412.1"
/db_xref="GI:16418965"
/translation="MARLAARDMGDTLLMPDHHGRETATATARRESDTITTFATGR
HYLEMRHLGTLSDAYLITGNGTRIRHSEBDVLRHRODLDQVADTVMMHAKMDTRASH
HVENDNMGFTQEIIPALQAHVYSGFRYQVIDIKSPAHQVTKLCPGSDHDLRLRI
QNLALBERAHLCSAVDCLEVLPLGCKSGALAVLSNHLGLSLADCAFGDANBRE
MLGSGVGRGLWGNAMPOLIALPHLSVIGHGNGQAVSHFLTHMLDNPILPYSPE"
complement(950..2010)
/gene="STM0458"
complement(950..2005)
/gene="STM0458"
/note="similar to E. coli cysteine synthase B, O-acetylserine sulfinhydrase B (AAC75474.1); Blastp hit to AAC75474.1 (303 aa), 26% identity in aa 6 - 207, 35% identity in aa 198 - 289"
/codon_start=1
/transl_table=1
/product="putative cysteine synthase/cystathionine beta-synthase"
/protein_id="AAL19413.1"
/db_xref="GI:16418966"
/translation="MMSSNRYKAININEINAHQGSADTHLRPLSAFPGIQLYKOE
STPLTGLKRLARSLFLYGLCNGMKEGTPIIASSGSTVIASEAFRLGLFLIAY
MPACTARKKIEIOIFYGCHHFVSACEIYVAASERLAEHLNGHMDQTVYERATDMR
GNNIADSIIPROMENELHPVRFIYMSAGTGTATGRYRCOGYDTQLVAVDPENS
VFLPYMODRSLSPVSGKIEGIRPVEPSPIPDVVDMLRPDASAVTAHMLLET
QLGRKVASGTCTNMGALQLAARKEAGETAIYTLCDSDRLIDTYIYHAWVSDHI
GDLTPWSAKIKKLIGD"
RBS /complement(2005..2010)
/gene="STM0458"
/note="putative RBS for STM0458; RegulonDB:STMS1H000803"
2107..2576
/gene="ybaO"
/note="synonym: STM0459"
2107..2112
/gene="ybaO"
/note="putative RBS for ybaO; RegulonDB:STMS1H000804"
2118..2576
/gene="ybaO"
/note="similar to E. coli putative LRP-like transcriptional regulator (AAC73550.1); Blastp hit to AAC73550.1 (181 aa), 95% identity in aa 30 - 181"
/codon_start=1
/transl_table=1
/product="putative transcriptional regulator (asnc family)"
/protein_id="AAL19414.1"
/db_xref="GI:16418967"
/translation="MLDKIDRLCLLOQDCTLHQLADAVNITTTFCWRLKRLLED
DGILLGRVALLDEKILGLTAFVLIKTHSHSSWMYCFYVSEMPYVIGFMMAGE
YDIIMRYVADMKRYDDPYKRLVNSVPLSVYTSFPAEQIKITTSLEPIE"
2604..4389
/gene="mdIA"
/note="synonym: STM0460"

RBS 2604..2609
/gene="mdIA"
/note="putative RBS for mdIA; RegulonDB:STMS1H000805"
2617..4389
/gene="mdIA"
/note="similar to E. coli ATP-binding component of a transport system (AAC73551.1); Blastp hit to AAC73551.1 (590 aa), 88% identity in aa 1 - 590"
/codon_start=1
/transl_table=1
/product="putative ABC superfamily (atp) transporter"
/protein_id="AAL19415.1"
/db_xref="GI:16418968"
/translation="NRLFAQLSWYFRFRERYGAVALMLMLQLIPKVGIVVD
GVTAQOFTPRIMMIGTALLIAVYVLLRYVRLVGLSAYOLAVELREYRQLSR
OHPEFYORHTGDLIARATNDVDRVYPAAGEGLTVDSLVMGCALVWSTOISML
TLIALPMPIMALMIKRYGDRLDYFKLQAARSSANDRQESLSTIRMKAGLEDR
QSSLFADADDTGKNNRVARIDARFDPTIYIAGNALIAISGSMVNVGSLTIGE
LTFPMYVIGIMVPMALAMFNIYERGSAAYSRIPAMLEAVVDGEEVPVAGOE
LTPAIRFCYPTTHPALENVNFRKPGOMLIGCPGAGKSTLSLQHPVTOGE
IRPHMDPLHLODSWRSRLAAYVSTPFLSPDSIANNIALGREATOEBIEOYARLS
VHEDILRLPGCVTOVGERGMVSGGOKORISIPALLNAETILLDDALSAYDGRE
HQLHNIRQMGSERTYIISARRSALTDAHEIIVMGHGVQGRDQLAOLQIGRYD
MYRYQLEALDPAERGEBAVNA"
4369..6163
/gene="mdIB"
/note="synonym: STM0461"
4369..4374
/gene="mdIA"
/note="putative RBS for mdIB; RegulonDB:STMS1H000806"
4382..6163
/gene="mdIB"
/note="similar to E. coli putative ATP-binding component of a transport system (AAC73552.1); Blastp hit to AAC73552.1 (593 aa), 88% identity in aa 1 - 593"
/codon_start=1
/transl_table=1
/product="putative ABC superfamily (atpmembrane) transporter"
/protein_id="AAL19416.1"
/db_xref="GI:16418969"
/translation="MRSFGQLMPTLRLRLAYGSPWRKPLSVAMMIMIAAAVSGPL
LIVYFDINWVAHRLPLGKVAALAAVYGLQFLAAGLHAQSLIFRNAAGVQSLST
DVAADLROGLSADFQVQGLISRYTNDREIDYLVVAVLRSALIGAMLVAM
FSLDMRMLPAILIIPVAVLTMIIYORSTPIYRRAVAYADINDGENEINGSVTA
OPROCARGERMGEASRSHYMAOTLRDGLFLRLSLFSALLICGLMLPSFSA
GTTVEGYLVAFIYLSRLNPLELLETTQSMLOQAAVYAGRVELMDPRQYSGSDR
PLQSGALDIDHLSFAYRDNLVQDITLSVPSRSFVALVGHTSKSTLASLNGYTP
LTQGEIRLDGREIASLNHRVLRQGVAMVQDDPVVMDTFLANVTLGRDVSBAQVWOL
ETVQLADLAGSLSDGHLTHLGEQNTLSVGOKLLALARVLVDAPOILLIDEATASID
SGTEOAIQCALAIRERTTLVIAHRLSTIVEADTIVLRGQAVERTHQQLAAGC
RYWOMVOCLVGBELAASVHEPGRPA"
6363..6713
/gene="glnK"
/note="synonym: STM0462"
6363..6368
/gene="glnK"
/note="putative RBS for glnK; RegulonDB:STMS1H000807"
6375..6713
/gene="glnK"
/note="nitrogen assimilation by glutamine synthetase; regulates GlnI (NRII) and GlnE (NRIa); similar to E. coli nitrogen regulatory protein P-II 2 (AAC73553.1); Blastp hit to AAC73553.1 (112 aa), 97% identity in aa 1 - 112"
/codon_start=1
/transl_table=1
/product="regulatory protein P-II 2"
/protein_id="AAL19417.1"
/db_xref="GI:16418970"
/translation="MKLVYIIRKFLYEDVREALSGIOGLTVTEVKGGRGXHAE
LIRGAETVNFLEPRKVIDAIDDDQLEBEVIDITISKAAYTKIGDGIKFAELMRVIRI
RTGEADEAAL"
6731..8031
gene


```

Db      3050 CAGAGGTCGATCTTACCTACGTTGATGAGCCCAATCCAGCAACGATTTATAGCGCTG 3109
Qy      234 ArgArgInlIeaIaGlUgLyLeuLeuAlaIaAspGlyValIleAlaValGlyIle 253
Db      3110 AAGCAATCCGAGGCAAGGAAATTTGATAGGAAAAAGCCACGACGATGAGCTTG 3169
Qy      254 LysPheAspAspGlyThrValTyProGlnuLysGlyArgLeuLeuPheAlaAspProVal 273
Db      3170 GTCATGAAACCGGTCAACCTATCCCTGAAAGTACCGTCGATCTCCGATGTACCC 3229
Qy      274 ValAsnGlnSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAspGlnAsnIle 293
Db      3230 GTTATGAAAGCACCGGCTCCATACCTACGCTGCTCTTCTTCCCTAACCCGCAACATACG 3289
Qy      294 LeuMetProGlyLeuTyValArgValLeuMetAspGlnAlaValAlaAspAsnAlaPhe 313
Db      3290 CTTTGGCGGGATGTTGTGTCGTCACGCGATTGATAGAGCGCTCCAACTGACGCCATT 3349
Qy      314 ValValProGlnGlnAlaValThrArgGlyAlaLysAsp-----ThrValMetIleVal 331
Db      3350 CTTATCCCGCAACAGGCGTTAGCCGACACCGCGGTGATGACACCGCTCATGTT 3409
Qy      332 AsnAlaGlnGlyMetGluProArgGluValThrValAlaGlnGlnGlyThrAsn 351
Db      3410 AACGATAAAGTCAGTTGAAGCGCCCTGTCTGTCAGTCAGGCGGATGGCGATAAA 3469
Qy      352 TrpIleValThrSerGlyLeuLysAspGlyAspLysValValValGlyGlyIleSerIle 371
Db      3470 TGGTTGATGTTGAGAGCATGAAATCTGGCGATCAAGTCATGTCACCGGCTG----- 3523
Qy      372 AlaGlyIleThrGlyAlaLysValThrProLysGluThrPalaSerSerGluAsnGln 391
Db      3524 -----CAAAAGCGCGTCCGGAAG----- 3544
Qy      392 AlaAlaAlaProGlnSerGlyValGlnThrAlaSerGlyAlaLysThrAlaSerGlu 410
Db      3545 -----CAGTTAAAGCCACTACGATACCCCGAGATGACTGATCGAAG 3589

RESULT 33
AE000405 11095 bp DNA linear BCT 01-DEC-2000
LOCUS Escherichia coli K12 MG1655 section 295 of 400 of the complete
DEFINITION genome.
ACCESSION AE000405.1 GI:1789659
VERSION AE000405.1
KEYWORDS
SOURCE Escherichia coli K12
ORGANISM Escherichia coli K12
REFERENCE 1 (bases 1 to 11095)
AUTHORS Blatterner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
Ballet,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
Manu,B. and Shao,Y.
TITLE The complete genome sequence of Escherichia coli K-12
JOURNAL Science 277 (5311), 1453-1474 (1997)
MEDLINE 97426617
PUBMED 9278503
REFERENCE 2 (bases 1 to 11095)
AUTHORS Blatterner,F.R.
TITLE Direct Submission
JOURNAL Submitted (16-JUN-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
REFERENCE 3 (bases 1 to 11095)
AUTHORS Blatterner,F.R.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459

```

```

REFERENCE 608-263-7459
AUTHORS 4 (bases 1 to 11095)
TITLE Plunkett,G. III.
JOURNAL Direct Submission
COMMENT Submitted (13-OCT-1998) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
This sequence was determined by the E. coli Genome Project at the
University of Wisconsin-Madison (Frederick R. Blattner, director).
Supported by NIH grants HG00301 and HG01428 (from the Human Genome
Project and NCHGR). The entire sequence was independently
determined from E. coli K12 strain MG1655. Predicted open reading
frames were determined using Genemark software, kindly supplied by
Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
have been correlated with genetic loci are being annotated with CG
Site Nos., unique ID nos. for the genes in the E. coli Genetic
Stock Center (CGSC) database at Yale University, kindly supplied by
Mary Berlyn. A public version of the database is accessible
(http://cgsc.biology.yale.edu). Annotation of the genome is an
ongoing task whose goal is to make the genome sequence more useful
by correlating it with other data. Comments to the authors are
appreciated. Updated information will be available at the E. coli
Genome Project's World Wide Web site
(http://www.genetics.wisc.edu). ** The E. coli K12 sequence and
its annotations are periodically updated; this is version M54. No
sequence changes. Annotation updates: updated gene identifications
and products; all new functional assignments courtesy of Monica
Riley; added promoters, protein binding sites, and repeated
sequences described in reference 1. The unique numeric identifiers
beginning with a lowercase 'b' assigned to each gene (protein- or
RNA-encoding) are now designated as gene synonyms instead of
labels. This should allow low them to be searched for in Entrez as gene
names.

FEATURES
source
location/Qualifiers
1..11095
/organism="Escherichia coli K12"
/mol_type="genomic DNA"
/strain="K12"
/sub_strain="MG1655"
/db_xref="taxon:83333"
7..35
/feature="factor sigma70; promoter fls; documented +1
at3407883"
76..1041
/feature="yhdG"
/feature="synonym: b3260"
76..1041
/feature="yhdG"
/function="putative enzyme; Not classified"
/feature="321, 100 pct identical amino acid sequence and
equal length to YHDG_ECOLI SM: P25717"
/codon_start=1
/product="putative dehydrogenase"
/protein_id="AAC76292.1"
/db_xref="GI:1789660"
/translation="NRIGQYQJLRNRIAPMACITRPPRTLCYEGAGITYSNMSS
NQVWESDSRLRMWIDPGRIVQIASDDEPMADARINVESAOIIDIMGCPA
KKYNRKLSALLOYDPDVVKSILTEVNVADVPTLKIRGMAPBEHRNCEBIAQYLED
CGIOALTIGRTPACTFNGEAEYDSIRAVQKVSIPVINGDITDPLKARAVADYGA
DALMTIGRAAGRPWIFREIOHYLDGELLPLPIAIVKRLCAHVEILDHPQAPAGY
RIARKHVSWYLOEHANPDQFRRTFNAINEDASQLELAEVFNFA"
946..972
/feature="yhdG"
/feature="factor sigma70; predicted +1 start at 3408820"
1067..1363
/feature="fls"
/feature="fls"
/feature="fls"
/feature="synonym: b3261"
1067..1363
/feature="fls"
/function="factor; DNA - replication, repair,
restriction/modification"
/feature="c98; CG Site No. 18328"

```

```

/promoter
/codon_start=1
/transl_table=1
/product="site-specific DNA inversion stimulation factor;
DNA-binding protein; a trans activator for transcription"
/protein_id="AAC76293.1"
/db_xref="GI:1789661"
/translation="MFEORVNSDVLTVSTVNSQDVQVQKPIRDSVKALKNYPAQLNG
QDVNDLYELVLAIEVEQPLDMWQYTRGNQTRALMGMINGITLRKLLKKGKMN"
1300..1328
/gene="fis"
/note="factor Sigma70; predicted +1 start at 3409176"
1443..2333
/gene="yhdv"
/note="synonym: b3262"
1443..2333
/gene="yhdv"
/function="putative enzyme; Not classified"
/note="0296; 100 pct identical amino acid sequence and
equal length to YHDV_ECOLI SW: P28638"
/codon_start=1
/transl_table=1
/product="putative methyltransferase"
/protein_id="AAC76294.1"
/db_xref="GI:1789662"
/translation="MTWKTGCEPTRFGNEAKTIHGDALELKKIPIAESVDLFIADPP
YNIKNPFDGLIEMKEDLFDLFEVLAECRVLKQSGMYINSTMPEPIDQCKR
LFTIKSRIVMSGDSGVQAKKGYSGMYEPIIMYKDAKNYFENDALIVEKTSQSA
LIDYKNDPDPYHOKVGNWDPFRVYLMDEVENHPQKPEALLRIILIASNPDP
LIDYKNDPDPYHOKVGNWDPFRVYLMDEVENHPQKPEALLRIILIASNPDP
TGNDLSKRSRLSEVPDLITK"
2349..2376
/note="factor Sigma70; predicted +1 start at 3410224"
2361..2389
/note="factor Sigma70; predicted +1 start at 3410237"
2417..2596
/gene="yhdv"
/note="synonym: b3263"
2417..2596
/gene="yhdv"
/function="orf; Unknown"
/note="059"
/codon_start=1
/transl_table=1
/product="orf, hypothetical protein"
/protein_id="AAC76295.1"
/db_xref="GI:1789663"
/translation="MIRKRYMWLVFAVVFPLFDLLMOWIELLATETDKCNMNSVP
LKLVCDELNFQDRN"
complement(2599..3261)
/gene="envr"
/note="synonym: b3264"
complement(2599..3261)
/gene="envr"
/function="putative regulator; Not classified"
/note="f220; 100 pct identical amino acid sequence and
equal length to ENVR_ECOLI SW: P31676"
/codon_start=1
/transl_table=1
/product="putative transcriptional regulator"
/protein_id="AAC76296.1"
/db_xref="GI:1789664"
/translation="MAKRTKAALKTROELIETALPAQHGSEKTTILNDADANYT
RGALVYHPEKNTOLFNEMWLOQPSLRELIQEHLAGLEHDPQOLRELIYGLQYIK
IIPQALILKILYHKCEPDEMLAGVIREKGFNPOTLRVLOACCOGCVANNLDD
VWIIIDGAFSGIVQNLMMNAGYDLYKQAPALVDNVLRMFPDENITKLHQNELS
VM"
3268..3288
/note="central position to predicted promoter: -226.5"
3268..3288
/note="central position to predicted site"
3268..3288
/note="central position to predicted promoter: -302.5"
3268..3288
/note="central position to predicted site"
3268..3288

```

```

/note="central position to predicted promoter: -318.5"
/bound_moiety="11v predicted site"
complement(3423..3451)
/note="factor Sigma70; predicted +1 start at 3411257"
3471..3499
/note="factor Sigma70; predicted +1 start at 3411347"
3547..3575
/note="factor Sigma70; predicted +1 start at 3411423"
3560..3591
/note="factor Sigma70; predicted +1 start at 3411439"
complement(3561..3578)
/note="central position to predicted promoter: -137"
/bound_moiety="Pn08 predicted site"
3660..4817
/gene="nacrE"
/note="synonym: b3265"
3660..4817
/gene="nacrE"
/function="membrane; inner membrane"
/note="0385; alternate name envC; 100 pct identical
aminoacid sequence and equal length to ACRE_ECOLI SW:
P24180"

```

```

Alignment Scores:
Pred. No.: 3,376-35 Length: 11095
Score: 750.50 Matches: 168
Percent Similarity: 60.65% Conservative: 74
Best local Similarity: 42.11% Mismatches: 122
Query Match: 37.17% Indels: 35
DB: 1 Gaps: 6

```

US-09-889-756a-2 (1-412) x AE000405 (1-11095)

```

Oy 17 AlavAlaLeuValLeuSerSerCyGlyGlyValAspAlaAlaGlnGlyGln 36
Db 3708 TCCGCGCTTTATATCCCGGTTGTAACGATAGGAGAAAGAAAGTCACTACGTGCA 3767
Oy 37 ProAlaGlyArgGluAlaProAlaProValAlaGlyValThrValHis--ProGln 55
Db 3768 CCGCAG-----GTTACCGTTCAATTGTAA 3794
Oy 56 ThrValAlaLeuThrVal-----GluLeuProGlyArgLeuGluSerLeuArgThrAla 73
Db 3795 ACCGCCCGCGTGAAGTTAAGTCACTGATCCAGCGCACCAATGCTTATCGATAGCC 3854
Oy 74 AspValArgAlaGlnValAlaGlyGlyLeuGluLeuLysArgLeuPheGlnGlySerTyr 93
Db 3855 GAAGTTCGCCACAGGTTAGCGGATGCTGATGCAATTCACCTGAAGCGCGAT 3914
Oy 94 ValArgAlaGlyGlnProLeuTyrGlnIleAspSerSerThrTyrGlnAlaAsnLeuGlu 113
Db 3915 GTGCACACAGCCAGCGCTCCGTGACCATGATCGATCCCGACATATACGCAATATATGAC 3974
Oy 114 SerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaLysAlaAspAlaAspLeu 133
Db 3975 AGCCCGAAAGCGCACTGCGGAAAGTGAAGCCGCCCCCATTCGCGCATTTGACCGTA 4034
Oy 134 AlaArgTyrLysProLeuValAlaAlaGluAlaValSerArgGlnGlyTyrAspAlaAla 153
Db 4035 AAAGTTACGTTCCGCTCGTGGGTACGAATATCATCCAGCAGGAGTACGACAGGCC 4094
Oy 154 ValThrAlaLysArgSerAlaGluAlaGlyValLysAlaAlaGlnAlaAlaIleLysSer 173
Db 4095 ATTGCTATGCTGCTACGCGCATGCGCGGTATGTCGCCAAAGCACAGTGAAGC 4154
Oy 174 AlaGlyIleAsnLeuAsnArgSerArgIleThrAlaProIleSerGlyPheIleGlyGln 193
Db 4155 GCTGCAATCAATCTTCTTATACCAAGTCACTGCGCAATTCGCGCATTCGCGCAA 4214
Oy 194 SerLysValSerGlyGlyThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIle 213
Db 4215 TCGACTGTACCGAAGCGCTTGTCTACTAATAGGCAAGCACTGACGACGTCTC 4274

```



```
QY 214 ArgGlnThrAspPrometCTYValAsnValThrGlnSerAlaSerGluValMetLysLeu 233
Db 4275 CAGCGCTCGATCTTACCTGATGATGACCCCAATCCAGCATTTTATGAGGCTG 4334
QY 234 ArgAsgGlnIleAlaGlnGlyLysLeuLeuAlaIleAspGlyValIleAlaValGlyLe 253
Db 4335 AACGATCCGTAGAGCAAGAAATTGCTTAAGAAAGCCACGACGACGATGAGTTG 4394
QY 254 LysPheAspAspGlyThrValIleProGlnLysGlyArgLeuLeuPheAlaAspProVal 273
Db 4395 GTCATGAAAGCGCTCAACCTATCCCTGAAAGGTAGCTGCATTTCTCCGATGTGAC 4454
QY 274 ValAsnGlnSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAspGlnAsnIle 293
Db 4455 GTTGATGAAGACACCGCTCCATACCTACGTCGTCTTCCCTAACCCGCAACATACG 4514
QY 294 LeuMetProGlyLeuThrValArgValLeuMetAspGlnValAlaValAspAsnAlaPhe 313
Db 4515 CTTTGGCCGGATGTTGTGCGTGACGAGTGTAGTGAAGCGCTCAACCTGACGCCATT 4574
QY 314 ValValProGlnGlnAlaValThrArgGlyAlaLysAsp-----ThrValMetIleVal 331
Db 4575 CTTATCCCGCAACAAGCGTTAGCGGACACCGCGTGTGATGCAACCGTCTGATTGTT 4634
QY 332 AsnAlaGlnGlyGlyMetGluProArgGluValThrValAlaGlnGlnGlyThrAsn 351
Db 4635 AACCATTAAGTCAAGTGAAGCGCGCTCTGTCCTGTCACATGCGATGGCGATTA 4694
QY 352 TrpIleValThrSerGlyLeuLysAspGlyAspLysValValValGlnGlyIleSerIle 371
Db 4695 TGGTGTATTAGTGAAGACTGAATCTGGCGATCAAGTCATGTCACGCGGCTG----- 4748
QY 372 AlaGlyIleThrGlyAlaLysLysValThrProLysGluTrpAlaSerSerGluAsnGln 391
Db 4749 -----CAAAAGCGGCTCGGAGAG----- 4769
QY 392 AlaAlaAlaProGlnSerGlyValGlnThrAlaSerGluAlaLysThrAlaSerGlu 410
Db 4770 -----CAGGTTAAAGCCACTACGATACCCCGCAGATGACTGATCGAAG 4814

RESULT 34
ECOM67_1
WPCOMMENT
Sequence split into 4 fragments LOCUS ECOM67 Accession U18997
Fragment Name Begin End
ECOM67_0 1 110000
ECOM67_1 100001 210000
ECOM67_2 200001 310000
ECOM67_3 300001 372438
Continuation (2 of 4) of ECOM67 from base 100001 (U18997 Escherichia coli K-12 chromosom

Alignment Scores:
Pred. No.: 6 77e-34 Length: 110000
Score: 750.50 Matches: 168
Percent Similarity: 60.65% Conservative: 74
Best Local Similarity: 42.11% Mismatches: 122
Query Match: 37.17% Indels: 35
DB: 1 Gaps: 6

US-09-889-756a-2 (1-412) x ECOM67_1 (1-110000)
QY 17 AlaValAlaLeuValLeuSerSerCysGlyLysGlyLysAspAlaIleGlnGlyGln 36
Db 94658 TCCGCGGCTTAAATCGCGGCTTGAAGATTAAGAGAGAGAAAGCTCAGTCGATGA 94717
QY 37 ProAlaGlyArgGluAlaProAlaProValValGlyValValThrValHis--ProGln 55
Db 94718 CCGGAG-----GTTACCGTTCATATTGTAA 94744

QY 56 ThrValAlaLeuThrVal-----GluLeuProGlyArgLeuGluSerLeuArgThrAla 73
Db 94745 ACGGCCCCGTTAAGATTAAAGACTGAATTACAGGCGGACCAATGCTTATCGATACGCC 94804
```

```
QY 74 AspValArgAlaGlnValGlyIleIleGlnLysArgLeuPheGlnGlnGlySerTyr 93
Db 94805 GAATGTGCGCCACAGATTAGCGGATGCTGATGATGCAATTCATGAGAGGACGAT 94864
QY 94 ValArgAlaGlnProLeuThrGlnIleAspSerSerThrTyrGluAlaAsnLeuGln 113
Db 94865 GTGCAAGAGGCGGCTCGTGTACCAATCATCCGCGACCTATACGGAATAATTATGAC 94924
QY 114 SerAlaArgAlaGlnAlaThrAlaGlnAlaThrLeuAlaLysAlaAspAlaAspLeu 133
Db 94925 AGCCGAAAGGCGAAGTGGGCAAAAGTGAAGCCGCCCGCATCGGACATTGAAGGTA 94984
QY 134 AlaArgTyrLysProLeuValAlaAlaGlnAlaValSerArgGlnLysArgAspAla 153
Db 94985 AAAGCTTACGTTCCGCTCGTGCGTACGAATATCATGACGACGAGTACGACAGGCC 95044
QY 154 ValThrAlaLysArgSerAlaGluAlaGlyLysAlaLysAlaAlaIleLysSer 173
Db 95045 ATTGCTGATGCTCTCAGCGCGATGCGCGGTGATTCGCGAAAGCCACAGTGAAAGC 95104
QY 174 AlaGlyIleAsnLeuAsnArgSerArgIleThrAlaProIleSerGlyPheIleGlyGln 193
Db 95105 GCTGCAATCAATCTTCTTATACCAAGTCACTGCGCAATTAAGCGAGCATGAGGCAA 95164
QY 194 SerLysValSerGlnGlyThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIle 213
Db 95165 TCGACTGTGACCGAAGCGCTCTTGTCAATAGGCGAAAGCATGAACTGAGCGACGCTC 95224
QY 214 ArgGlnThrAspPrometCTYValAsnValThrGlnSerAlaSerGluValMetLysLeu 233
Db 95225 CAGCAGCTCATCTTACTGATGATGATGATGATGATGATGATGATGATGATGATGATG 95284
QY 234 ArgArgGlnIleAlaGlnGlyLysLeuLeuAlaAlaAspGlyValIleAlaValGlyLe 253
Db 95285 AAGCAATCCGTAGACAGAGAAATTGCTAAGAGAAAGCCACACGACCAAGTGAAGTGG 95344
QY 254 LysPheAspAspGlyThrValIleProGlnLysGlyArgLeuLeuPheAlaAspProVal 273
Db 95345 GTCATGAAAGCGTCAACCTATCCCTGAAAGTACGCTGCAATTCCTCGATGTAC 95404
QY 274 ValAsnGlnSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAspGlnAsnIle 293
Db 95405 GTTGATGAAGACACCGCTCCATACCTACGTCGTCTTCCCTAACCCGCAACATACG 95464
QY 294 LeuMetProGlyLeuThrValArgValLeuMetAspGlnValAlaValAspAsnAlaPhe 313
Db 95465 CTTTGGCCGGATGTTGTGCGTGACGAGTGTGAAGCGCTCAACCTGACGCCATT 95524
QY 314 ValValProGlnGlnAlaValThrArgGlyAlaLysAsp-----ThrValMetIleVal 331
Db 95525 CTTATCCCGCAACAAGCGCTTACCGCACCGCGTGTGATGCAACCGTCTGATTGTT 95584
QY 332 AsnAlaGlnGlyLysMetGluProArgGluValThrValAlaGlnGlnGlyThrAsn 351
Db 95585 AACGATTAAGTCAAGTGAAGCGCGCTGCTGTTGCGAGTCAAGCGATGAGCATTA 95644
QY 352 TrpIleValThrSerGlyLeuLysAspGlyAspLysValValValGlnGlyIleSerIle 371
Db 95645 TGGTGTATTAGTGAAGACTGAATCTGGCGATCAAGTCAATTCGACGCGCTG----- 95698
QY 372 AlaGlyIleThrGlyAlaLysLysValThrProLysGluTrpAlaSerSerGluAsnGln 391
Db 95699 -----CAAAAGCGGCTCGGAGAG----- 95719

QY 392 AlaAlaAlaProGlnSerGlyValGlnThrAlaSerGluAlaLysThrAlaSerGlu 410
Db 95720 -----CAGGTTAAAGCCACTACGATACCCCGCAGATACCTGATGAG 95764

RESULT 35
AE005554 11807 bp DNA linear BCT 21-MAR-2001
LOCUS Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 173
DEFINITION of 290.
```

```

ACCESSION      AE005554 AE005174
VERSION        AE005554.1 GI:12517891
KEYWORDS
SOURCE
ORGANISM       Escherichia coli O157:H7 EDL933
                Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
                Enterobacteriaceae; Escherichia.
REFERENCE      1 (bases 1 to 11807)
AUTHORS        Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
                Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
                Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
                Grobbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
                Apodaca,J., Anantharaman,T.S., Lin,Y., Yen,G., Schwartz,D.C.,
                Welch,R.A. and Blattner,F.R.
                Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
                Nature 409 (6819), 529-533 (2001)
                21074935
TITLE          JOURNAL
MEDLINE
PUBMED         11206551
REFERENCE      2 (bases 1 to 11807)
AUTHORS        Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
                Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
                Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
                Grobbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
                Apodaca,J., Anantharaman,T.S., Lin,Y., Yen,G., Schwartz,D.C.,
                Welch,R.A. and Blattner,F.R.
                Direct Submission
                Submitted (22-OCT-2000) Laboratory of Genetics, University of
                Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
SOURCE
                location=Qualifiers
                1..11807
                /organism="Escherichia coli O157:H7 EDL933"
                /mol_type="genomic DNA"
                /strain="EDL933"
                /serotype="O157:H7"
                /db_xref="taxon:155864"
                /note="enterohaemorrhagic"
                70..960
                /gene="yhndJ"
                /note="synonym: Z4622"
                70..960
                /gene="yhndJ"
                /function="putative enzyme; Not classified"
                /note="Residues 1 to 296 of 296 are 98.98 pct identical to
                residues 1 to 296 of 296 from Escherichia coli K-12 Strain
                MG1655: B3262"
                /codon_start=1
                /transl_table=11
                /product="putative methyltransferase"
                /protein_id="AAG58390.1"
                /db_xref="GI:12517892"
                /translation="MTMRGCEPTFRFGNEAKTIIHGDALAEKKLPESVDLIPADPP
                YNIGKMPGLIEAKKEDLFTDMLLEVAIEGRVAKKQSGWYINNSTEMPTIDLOCKR
                LPTIKSRIVMSDSSGVQAKKGYGSMYEPITLMMYKDAKNTYFNDALIVEAKTSQRA
                LIDRKNPQPPYNHQKPGVNWDPFRVRLYIDREYENHTQKPEALKKTIILASNPGD
                IVLDPFASFTGAVALASGRKFTGIEINSEYIKWGLRRLLDVASHYSABELAKVKKRC
                TGNLSKRSRLSEVDPDLITK"
                1044..1223
                /gene="yhndJ"
                /note="synonym: Z4623"
                1044..1223
                /gene="yhndJ"
                /function="orf; Unknown function"
                /note="Residues 1 to 59 of 59 are 100.00 pct identical to
                residues 1 to 59 of 59 from Escherichia coli K-12 Strain
                MG1655: B3263"
                /codon_start=1
                /transl_table=11
                /product="orf, hypothetical protein"
                /protein_id="AAG58391.1"
                /db_xref="GI:12517893"
                /translation="WIKRYMWLVVPAVAVFLPDTLLMQWIELLATETDKCKNNVNP
                LKLVNCELFQDRM"
                complement(1226..1888)
gene

```

```

CDS
                /gene="envr"
                /note="synonym: Z4624"
                complement(1226..1888)
                /gene="envr"
                /function="putative regulator; Not classified"
                /note="Residues 1 to 220 of 220 are 100.00 pct identical
                to residues 1 to 220 of 220 from Escherichia coli K-12
                Strain MG1655: B3264"
                /codon_start=1
                /transl_table=11
                /product="putative transcriptional regulator"
                /protein_id="AAG58392.1"
                /db_xref="GI:12517894"
                /translation="MAKRTAEELKTRQELIETAAQAPAGVSKTTLNDIADAANT
                RGAIVHFERKTLQFNEMLQPSLELIEHETAGIEHPFOQLREKLIYGLQYIAK
                IPRQALLKTIYHKCEFNDEMLAEGVIREKGNPOTLRVLAQCOQCGVANNLDD
                VMMIIDGASGIVQNMVLMNAGVDLYKQAPALVDNVLNFMFMDENTITKLIHQNELS
                VM"
                2287..3444
                /gene="acrE"
                /note="synonym: Z4625"
                2287..3444
                /gene="acrE"
                /function="membrane; Cell envelop: Inner membrane"
                /note="Residues 1 to 385 of 385 are 99.74 pct identical to
                residues 1 to 385 of 385 from Escherichia coli K-12 Strain
                MG1655: B3265"
                /codon_start=1
                /transl_table=11
                /product="transmembrane protein affects septum formation
                and cell membrane permeability"
                /protein_id="AAG58393.1"
                /db_xref="GI:12517895"
                /translation="MTKHARFLLPSRLISALLIAGCNKGKHAHVGRQYVHVH
                KTAPEKTELPGRNTAAYRAEVRPOVSGIVLARNPREGSDVQAGSLYQIDATPYA
                NIDSAKGEILAKSSAAPAIATLVKRYVPLVGTIYSOEEDLADARQDAVATLAK
                ATVESARINLAYTKTAPISGRIGSVTEGALVTNGQTELAIVQOLDPIYDVYQS
                SNDPMRLKQSEQGNLKHENATNVEIMENGGQTPYIKTGLQPSDVLVDSSTILR
                AVFNPQHTLLPGMFVARIDEGVDPALIIPOGVSRTPRGATVLIIVDKSQVEVR
                PVVASQAGIDKMLISGLKSGQDVIVSGIQAKARPGQVAKTTPTPADTAK"
                3456..5342
                /gene="acrF_1"
                /note="synonym: Z4626"
                3456..5342
                /gene="acrF_1"
                /function="putative membrane; Protection responses:
                Drug/analog sensitivity"
                /note="Residues 1 to 620 of 628 are 99.19 pct identical to
                residues 1 to 620 of 1034 from Escherichia coli K-12
                Strain MG1655: B3266"
                /codon_start=1
                /transl_table=11
                /product="N-terminal fragment of integral transmembrane
                protein involved with acridine resistance"
                /protein_id="AAG58394.1"
                /db_xref="GI:12517896"
                /translation="MANFIRRPDIPAVLAILMMAGALILQLPVQVPTIAPPAS
                VSAIVPGADQGVDTVTVYIRBONNGINDLMWSSSDAGSVTLITFROSGDPI
                AQVQVQKQLATPLPQEVQOGISYKSSSYLWAGVSDNPDTQDISDYAS
                NVKDTLSRLNSVDVQDFGAQYAMRIWLDLNLKTKLTVDAVINOLKVNQAQIAAQ
                LGGTALPAGQGLNADSIIAQTRLNKPEBGFVTIRVSDGSDVRLKQVARELAGENPQ
                VIARINKPAAAGIKIATGANALDPAKAKLAELQEPFGQMKVLVYDTPFVQ
                LSIHEVYKTLFEALIMLVLMVFLQMRATLPTTAVPVLLGFTAILAAGYSINT
                LTMFGWALAGLIVDAIVVENVERVMEKDKLPKREATKSKSQIGALVGLAMYS
                AVFIEMAFEGSGTAIRQPSITVSAMALSVVALILVLPALCATLKPFSFHEK
                GEGFMENFTFDHSVNHVYNSVGKIGSTRIYLLIYALVAGVAVFLRLPSFLEPE
                DGVFTLMQLPAGATQERTQKVLIDVTVYLLKMERANESVFTVNGFSFGCAPOP
                HVT"
                5317..5369
                /note="O-island #129; Region of the EDL933 chromosome not
                homologous to E. coli K-12 MG1655"
                5373..6605
                /gene="acrF_2"
gene

```

misc_feature

gene

```

CDS
/feature="synonym: Z4627"
5373..6605
/feature="acrf 2"
/function="putative membrane; Protection responses:
Drug/analogue sensitivity"
/note="Residues 1 to 410 of 410 are 99.02 pct identical to
residues 625 to 1034 of 1034 from Escherichia coli K-12
Strain MG1655: B3266"
/codon_start=1
/translation=1
/product="C-terminal fragment of integral transmembrane
protein involved with acidine resistance"
/protein_id="AAG58395.1"
/db_xref="GI:12517897"
/translation="MAFVSLKPMWEERNDENSAEAVIRAKMELGKIRDFVLPNMP
AIVELGTAGFDEFLIDKAGLDALTOARNOLLGMAOHPSLVSRPGLDTPQF
KLEVDOKQALGVSLSDINQITSTALGCTYVNDPDRGKRYVVOAARKPMLEED
VDLTVRSTNGENVPPSAPFTTSHWYVGSPLERYNGLPSEITGEAPGTSDDNAL
MENLAKLPAGIGVDTWMSYERLSGNQAPALVAISFVVFCLALVESMSIPVS
MLVPLGIVGLVLAATLEFNQKNDVYFWVGLTTIGLSAKNAIIVERAKDLMKEGK
VVEATMAVAMRLRPILMTSLAFLGVLPLAISNGASGAQNAIVGIGMGMVSATLL
AIFPVXFEFVIRCFKG"
6858..7079
/feature="synonym: Z4628"
/feature="yhdV"
/feature="yhdV"
/function="orf; Unknown function"
/note="Residues 1 to 73 of 73 are 100.00 pct identical to
residues 1 to 73 of 73 from Escherichia coli K-12 Strain
MG1655: B3267"
/codon_start=1
/translation=1
/product="orf, hypothetical protein"
/protein_id="AAG58396.1"
/db_xref="GI:12517898"
/translation="MKRLIPVALITLALACAHSPCVPYDDOGLVHTNTCKGTT
QDNWETAGAIAGCAAAVAGLTWGIATLSK"
7618..8535
/feature="yhdW"
/note="synonym: Z4629"
/feature="yhdW"
/function="putative transport; Not classified"
/note="Residues 1 to 305 of 305 are 99.67 pct identical to
residues 1 to 305 of 305 from Escherichia coli K-12 Strain
MG1655: B3268"
/codon_start=1
/translation=1
/product="putative periplasmic binding transport protein"
/protein_id="AAG58397.1"

Alignment Scores:
Pred. No.: 4.2e-35 Length: 11807
Score: 749.50 Matches: 168
Percent Similarity: 60.65% Conservative: 74
Best Local Similarity: 42.11% Mismatches: 122
Query Match: 37.12% Indels: 35
DB: 1 Gaps: 6

US-09-889-756a-2 (1-412) x AE005554 (1-11807)
Qy 17 AlaValAlaLeuValLeuSerSerCysGlyLysGlyValAspAlaIaIaGlnGlyGln 36
Db 2335 TCCGCGCGCTTAATCCGCGGTGTTACGATPAGGAGAGAAAGCTCACTCGGTGAA 2394
Qy 37 ProAlaGlyArgGlnAlaProAlaProValValGlyValValThrValHis--ProGln 55
Db 2395 CCGGAG-----GTTACCGTTCATTATTGTA 2421
Qy 56 ThrValAlaLeuThrVal-----GlnLeuProGlyValArgLeuGlnSerLeuAlaGlnAla 73
Db 2422 ACGGCGCGCTTAAGAACTTAAGACTTAACGAGCGCGCACCAATGCTTATCGTATAGCC 2481

```

```

Qy 74 AspValArgAlaGlnValGlyGlyIleIleGlnLysArgLeuPheGlnGlySerTyr 93
Db 2482 GAAGTTCGCCACAGGTTAGCGGATCGTACTGAATCGCAATTTCACTGAAGCAGCAT 2541
Qy 94 ValArgAlaGlyGlnProLeuTyrGlnIleAspSerSerThrTyrGlnAlaAsnLeuGln 113
Db 2542 GTCCAGACAGGCGCAGTCCGTGTCAGATCGATCCCGCAGCTATCCAGGCAAAATTATGAC 2601
Qy 114 SerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaLysAlaAspLeu 133
Db 2602 AGCGCAAAAGGCAACTGCGCAAAAGTGAAGCCGCCGCCATTCGCGCATTTGACGGTA 2661
Qy 134 AlaArgTyrLysProLeuValAlaAlaGlnAlaValSerArgGlnLysTyrAspAlaVal 153
Db 2662 AAACGTTAGCTTCGCTCGTGGGTAGCAAAATACATCAGCAGCAGAGTATCGACAGGCC 2721
Qy 154 ValThrAlaLysArgSerAlaGlnAlaGlyValLysAlaAlaGlnAlaAlaIleLysSer 173
Db 2722 ATTGCTATGCTCGCAGGCGCATGCGCGCTGATTTGCGCAAAACGACGTCGAAGC 2781
Qy 174 AlaGlyIleAsnLeuAsnArgSerArgIleThrAlaProIleSerGlyPheIleGln 193
Db 2782 GCTCGATCATCTTGTCTTATACCAAGTACTGCGCAATTAGCGGACGTATCGCAAA 2841
Qy 194 SerLysValSerGlnGlyThrLeuLeuAsnAlaGlyAspThrValLeuAlaThrIle 213
Db 2842 TCGACGTGACCGCAAGCGCTCTTGCATCAATAGGCAAAAGCATGAACTGCGCAACTGTC 2901
Qy 214 ArgGlnThrAsnProMetTyrValAsnValThrGlnSerAlaSerGlnValMetLysLeu 233
Db 2902 CACGACCTGCATCTTATCTTATGTTGATGAGACCAATCCAGCAACGATTTTATAGGCTG 2961
Qy 234 ArgArgGlnIleAlaGlnGlyLysLeuLeuAlaAlaAspGlyValIleAlaValGlyIle 253
Db 2962 AACGACGTCCGTAGAGAGAAATTTGCATPAGAAAGCCCAACGCAACGTAGACTTG 3021
Qy 254 LysPheAspAspGlyThrValTyrProGlnLysGlyValArgLeuLeuPheAlaAspProVal 273
Db 3022 GTCATGGAACACGCTCAACCTTATCCCTGAAAGTACGCTGCAATTCCTCGATGAC 3081
Qy 274 ValAsnGlnSerThrGlnIleThrLeuArgAlaAlaValProAsnAspGlnAsnIle 293
Db 3082 GTTGATGAAGACACCGCTCCATPACCTGATGCTGCTTCCCTAACCCGCAACATACG 3141
Qy 294 LeuMetProGlnLeuTyrValArgValLeuMetAspGlnValAlaValAspAsnAlaPhe 313
Db 3142 CTTTCCCGGCTATGTTTGTGCGTGAACGATGATGAAGCGCTCCAACTGACGCCATT 3201
Qy 314 ValValProGlnGlnAlaValThrArgGlyAlaLysAsp-----ThrValMetIleVal 331
Db 3202 CTCATCCCGCAACAGCGCTTACCGCACACCGCGTGTATGCAACCGTGCATGTTGTT 3261
Qy 332 AsnAlaGlnGlyLysMetGluProArgGluValThrValAlaGlnGlnGlnGlyThrAsn 351
Db 3262 AACGATAAAGTACAGGTGGAAGTCCCGCTGCTGTTGCCAGTCAGCGCATGCGATAAA 3321
Qy 352 TrpIleValThrSerGlyLeuLysAspGlyAspLysValValValGlnGlyIleSerIle 371
Db 3322 TGGTATATTAGCAAGAGTCTGAATCTGGGATCAAGTCACTTGTCAAGCGGCTCG----- 3375
Qy 372 AlaGlyIleThrGlyLysLysValThrProLysGlyThrAlaSerSerGluAsnGln 391
Db 3376 -----CAAAAGCGCGCTCCGGAGAG----- 3396
Qy 392 AlaAlaAlaProGlnSerGlyValGlnThrAlaSerGlnAlaLysThrAlaSerGln 410
Db 3397 -----CAGTTAAAGCACTACGAGTAAACCCCGCAGATATCTGATCGAAG 3441
RESULT 36
AP002564 307962 bp DNA linear BCT 07-MAR-2001
LOCUS Escherichia coli O157:H7 DNA, complete genome, section 15/20.
DEFINITION

```

```

ACCESSION      AP002564 BA000007
VERSION        AP002564.1 GI:13363382
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS        Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
                Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
                Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
                Sasakawa,C. and Shinagawa,H.
                Complete nucleotide sequence of the prophage VT2-Sakai carrying the
                verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
                derived from the Sakai outbreak
                Genes Genet. Syst. 74 (5), 227-239 (1999)
JOURNAL        20198780
MEDLINE        10734605
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL        20557356
MEDLINE        11108008
PUBMED
REFERENCE
AUTHORS        Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
                Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
                Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
                Shinagawa,H.
                Complete nucleotide sequence of the prophage VT1-Sakai carrying the
                Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
                O157:H7 strain derived from the Sakai outbreak
                Gene 258 (1-2), 127-139 (2000).
JOURNAL        1111050
MEDLINE
PUBMED
REFERENCE
AUTHORS        Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
                Yokoyama,K., Han,C.-G., Ohnishi,M., Nakayama,K., Murata,T.,
                Tanaka,M., Tohe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
                Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
                Shinagawa,H.
                Complete genome sequence of enterohemorrhagic Escherichia coli
                O157:H7 and genomic comparison with a laboratory strain K-12
                DNA Res. 8 (1), 11-22 (2001)
JOURNAL        21156231
MEDLINE
PUBMED
REFERENCE
AUTHORS        Hayashi,T., Makino,K., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
                Hayashi,T.
                Direct Submission
                Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
                Information Research Center, 3-1, Yamadaoka, Suita, Osaka 565-0871,
                Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
                URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
                Fax:81-6-6879-2047)
                genome project.
                Location/Qualifiers
                1. 307962
                /organism="Escherichia coli O157:H7"
                /mol_type="genomic DNA"
                /strain="O157:H7"
                /sub_strain="RIMD 0509952"
                /db_xref="taxon:83334"
                72..653
                /gene="ECs3910"
                72..653
                /gene="ECs3910"
                /note="similar to MDAB_ECOLI gi|1789406 percent identity
                100 in 193 aa (Conserved in E.coli K-12)"

```


DEFINITION	Stenotrophomonas maltophilia smeD, smeE and smeF genes.
ACCESSION	AJ252200
VERSION	AJ252200.1 GI:11071582
KEYWORDS	outer membrane protein; putative membrane fusion protein; putative RND protein; smeD gene; smeE gene; smeF gene.
SOURCE	Stenotrophomonas maltophilia
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales; Xanthomonadaceae; Stenotrophomonas.
REFERENCE	1
AUTHORS	Alonso, A. and Martinez, J.J.
TITLE	Cloning and characterization of SmeDEF, a novel multidrug efflux pump from Stenotrophomonas maltophilia
JOURNAL	Antimicrob. Agents Chemother. 44 (11), 3079-3086 (2000)
MEDLINE	20493115
PUBMED	11036026
REFERENCE	2 (bases 1 to 5914)
AUTHORS	Martinez, J.
TITLE	Direct Submission
JOURNAL	Submitted (05-JAN-2000) Martinez, J., Departamento de Biocietologia Microbiana, Centro Nacional de Biocietologia CSIC, Campus Universidad Autonoma, Cantoblanco, Madrid, 28049, SPAIN
FEATURES	Location/Qualifiers
SOURCE	1..5914
	/organism="Stenotrophomonas maltophilia"
	/mol_type="genomic DNA"
	/isolate="D457R"
	/db_xref="taxon:40324"
	/note="D457R is a multiresistant mutant obtained by single-step selection on tetracycline of D457"
	82..1266
	/gene="smeD"
	82..1266
	/gene="smeD"
	/function="component of a multidrug efflux system"
	/codon_start=1
	/transl_table=11
	/product="putative membrane fusion protein"
	/protein_id="CAC14594.1"
	/db_xref="GI:11071583"
	/translation="MLSRIRPPLSLIAIAFVAACGGPOAPRPGPDVTVTLKSE
	TVGLTRLEPGRNTAFVAEVRPQNGVIYAKRLPREGGVKAGEPLXQLDASVRAQAN
	NARQQLAREATIANARLSAKRITELAKVDVVSQDLENVAIPAQKADADGAKKASI
	DAANVTGGYARITAIPIASGRIGKSVTQGLVASQANALVAVQQLDPIYDLTQSSAE
	LLODRELIAAGRLQDNQTLPVSLIMEDSTEEHGTLEFSEVSDPTTSGTGLKVKVD
	NPDGLMGMYVRAIVIGGVSADALVIMOGIARPKDDTAMVVGKQNKREVPVKVK
	SRVYDGKMLVEDGLKAGDKYIVEGQIKGPEMPVKATIKGDAAPKPAALAAPADG
	AK"
	1279..4401
	/gene="smeE"
	1279..4401
	/gene="smeE"
	/function="component of a multidrug efflux system"
	/codon_start=1
	/transl_table=11
	/product="putative RND protein"
	/protein_id="CAC14595.1"
	/db_xref="GI:11071584"
	/translation="MARFPIRPIFAWYIATITMAGLAFKLPVSNYPNVAFAVE
	ISATYPSAKVDESVTQIITBQNKGDGLIYSSNSSNGQATITTFESGTPDIT
	AOVQNKQLQAMPLLPQEVORQGINVAKSSGFNALAFSENGSMANDIAIVQSG
	NVVDLSRVPGVNTQVFGKRYAMRIWDPKNLHTYGSVEYVATIAQNAQVALIGQ
	LGDAISVGGQQLNATINQSRLOTEPECPRIIVISGAODGALRLGDVAVRLGAEVD
	FVTRNGQPSGLAVTLTGANLADPTAGVADALVADMGGFPAGIKAIPIYDTPFVR
	VSIKGVNTLEIATVIVFWMYLPIQNRATLPIPIAVPVILGPGVLMGSSVMV
	LTNFMVLAIGLVDDAIVVENVBRINSERGLSPLEKTRSKMOITALTGIGLIVLS
	AVFPMAFMSGSTGVITROFSATIVSAALSVLAIVTTPALCATMLKPLKGEHVA
	HRLGAFREFGNSGTRFRTSISYQRGVGIHRRPFRMGIVAAFLVNGVLVFLPSS
	FLPNDDQVLMALVOAPVAGATQERTLSIALLEENGLDNEADVDSVSGFSPAGG
	GONMGAEVULKDMSEKADNGWMLTGRAMAAIGQIDAFIPAPPAPAIPELGTA
	YTFPIKNSGCGHVALVARNQLGLAGSKLNNVRNGEDDTPQPIIDAKATG
	LGSLDILNGTLIAAMGSSYITDFFDRKRVKRVQADQARFMPEBDLMSVKNK
	EMWPSAATKMGWDGSPRLRINYGVSAWLGEPAPGVASGDMAELBOLAKLPAG

[illegible]


```

Qy 174 AlaglyleasleuAsnArgSerArgIleThraIaProIleSergIlyPheIleGlyGln 193
Db 215742 GCTCGCATCAATCTCGCTTAACCAAGTAGCTCGCAATTAACGGACGATCGCAAA 215801
Qy 194 SerIysIalSerIugIyThrLeuAsnIaIaGlyAspIlyThrIaValLeuIleThIle 213
Db 215802 TCGACCGTGAACCGGAGCTCTTGTCATCAATAGGCAACAGCTGAACCTGGCACTGTC 215861
Qy 214 ArgGlnThrAsnPrometIyValAsnValThrgInSeralAsergIuValMetIyleu 233
Db 215862 CAGAGCTCATCTTATCTACGTGATGTGACCCCAATCCACACACATTTATAGGCTG 215921
Qy 234 ArgArgGlnIlealagIlyIyLeuLeuIaIaIaAspIyValIlealValGlyIle 253
Db 215922 AAGCAATCCGTAAGCAAGGAAATTTGATAGGAAACCGCACAGCAAGTGAAGTGTG 215981
Qy 254 LysPheAspAspIyThraIyValIyProGlnIyIyIyIyIyIyIyIyIyIyIyIyIy 273
Db 215982 GTCATGAGAACGGTCAAACTTATCCCTGAAAGTACGCTGCAATTCCTCGATGTGACC 216041
Qy 274 ValAsnGlnSerThrgIyGlnIleThrLeuArgIaIaIaValProAsnAspGlnAsnIle 293
Db 216042 GTTGATGAAGACACCGGCTCATTAACCTTACGTCTGCTTCTTCCATACCGCAACATACG 216101
Qy 294 LeuMetProGlyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 313
Db 216102 CTTTTCGCGGTATGTGTGTCGCGCACGATGATGAAAGCGTCCAACTGCGCATTCATT 216161
Qy 314 ValValIProGlnIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 331
Db 216162 CTTATCCCGCAACAGCGGTGACCGCACACAGCTGTGTGACACCGCTGCTGATGTGT 216221
Qy 332 AsnIaIaGlnIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 351
Db 216222 AACCAATTAAGTCAAGTGAAGCCGCCCTCTCTGTCACATCAGCGCATTCGGCATAAA 216281
Qy 352 TrpIleValThrSergIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 371
Db 216282 TGGTTGATTAAGCGAAGCTGAATCTGGCATCAAGTCATGTTCAGCGGCTG----- 216335
Qy 372 AlaGlyIleThrGlyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 391
Db 216336 -----CAAAAAGCGCTGCGGAGAG----- 216356
Qy 392 AlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 410
Db 216357 -----CAGTTAAAGCCATACCATACCCCGCATCTGCGATCGAAG 216401

RESULT 39
AE011925/c 12478 bp DNA linear BCT 29-MAY-2002
LOCUS Xanthomonas axonopodis pv. citri str. 306, section 303 of 469 of
DEFINITION the complete genome.
ACCESSION AE011925 AE008923
VERSION AE011925.1 GI:21109129
KEYWORDS
SOURCE Xanthomonas axonopodis pv. citri str. 306
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
Xanthomonadaceae; Xanthomonas.
REFERENCE
1 (bases 1 to 12478)
da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R.,
Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida
Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C.,
Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardoso,J.,
Chambergo,F., Ciapina,L.P., Cicarelli,R.M.B., Coutinho,L.L.,
Cursino-Santos,J.R., El-Dorry,H., Faria,J.B., Ferreira,A.J.S.,
Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C.,
Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,I.T., Leite
Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A.,
Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Meidanis,J.,
Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M.,
Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A.,

```

```

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinoia,L.A.F.,
Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos
Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and
Kittajima,J.P.
Comparison of the genomes of two Xanthomonas pathogens with
differing host specificities
Nature 417 (6887), 459-463 (2002)
2 (bases 1 to 12478)
da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R.,
Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida
Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C.,
Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardoso,J.,
Chambergo,F., Ciapina,L.P., Cicarelli,R.M.B., Coutinho,L.L.,
Cursino-Santos,J.R., El-Dorry,H., Faria,J.B., Ferreira,A.J.S.,
Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C.,
Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,I.T., Leite
Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A.,
Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Meidanis,J.,
Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M.,
Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A.,
Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinoia,L.A.F.,
Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos
Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and
Kittajima,J.P.
Direct Submission
Submitted (28-NOV-2001) Departamento de Biologia, Universidade de
Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900,
Brazil
FEATURES
source
location/Qualifiers
1. 12478
/organism="Xanthomonas axonopodis pv. citri str. 306"
/mol_type="genomic DNA"
/strain="306"
/db_xref="taxon:190486"
/note="pathovar: citri"
complement (72. 1103)
/gene="moca"
/note="synonym: XAC2835"
complement (72. 1103)
/gene="moca"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/transl_table=1
/codon_start=1
/product="oxidoreductase"
/protein_id="AA037680.1"
/db_xref="GI:21109130"
/translacion="METRLGRSGFVKPVVLGAGTGGKPLFSAMGDTDVAAARM
IDICLAGINLPPTAVVSDGASEDILGALOGRRDVLITSTGRLGSGPDAGAS
RRRLIAYENCRLRLTDYIDILQIAPDMTMEETLSTDLGVARGKRYIGASNY
SGQILKMSQAVADPQHMNRVNAVQITYSIGRVEWELMPLGIDOGIGAIVSPLNG
RLTGKLRGQPLPQHSRLHATAAGAIIDARLFDIVLADAIAAETGRVSPQAIWM
LQRPVTVTLIGARDEVQLRENIAGVWSLTGDOBARILDQSAVEPPYPYPRWQ
FAERSPLPL"
complement (1260. 1403)
/gene="XAC2836"
complement (1260. 1403)
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/transl_table=1
/codon_start=1
/product="WFS transporter"
/protein_id="AA037681.1"
/db_xref="GI:21109131"
/translacion="MTATQAGVATPWAALLSAVGIGALMSVQIGRRRAAPLNTC
QMG"
complement (1476. 2477)
/gene="arab"
/note="synonym: XAC2837"
complement (1476. 2477)
gene
CDS

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2003, 14:09:43 ; Search time 25 Seconds

(without alignments)
2262.764 Million cell updates/sec

Title: US-09-889-756A-2

Perfect score: 412
Sequence: 1 MAFAFKAMRAAALAAVAL.....AAPSGVOTASEAKTASEAE 412

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 513375 seqs, 137303645 residues

Word size : 0

Total number of hits satisfying chosen parameters: 513375

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : Published Applications_Aa.*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCT_US_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10	2.4	124	US-10-012-542-398	Sequence 398, App
2	10	2.4	222	US-09-890-688-154	Sequence 154, App
3	8	1.9	102	US-10-156-761-10004	Sequence 10034, A
4	8	1.9	181	US-10-080-170-577	Sequence 577, App
5	8	1.9	194	US-10-012-542-397	Sequence 397, App
6	8	1.9	478	US-09-813-555-1	Sequence 1, Appli
7	8	1.9	478	US-09-813-555-2	Sequence 2, Appli
8	8	1.9	478	US-09-813-555-3	Sequence 3, Appli
9	8	1.9	478	US-09-943-123-6	Sequence 6, Appli
10	8	1.9	478	US-09-943-123-7	Sequence 7, Appli
11	8	1.9	478	US-09-943-123-12	Sequence 12, Appli
12	8	1.9	478	US-09-943-123-13	Sequence 13, Appli
13	8	1.9	480	US-09-943-123-16	Sequence 16, Appli
14	8	1.9	480	US-09-943-123-17	Sequence 17, Appli
15	8	1.9	500	US-10-106-698-4765	Sequence 4765, App

16	8	1.9	891	US-09-815-242-10147	Sequence 10147, A
17	8	1.9	892	US-09-815-242-13843	Sequence 13843, A
18	7	1.7	17	US-09-864-761-36687	Sequence 36687, A
19	7	1.7	24	US-09-983-802-347	Sequence 347, App
20	7	1.7	37	US-09-864-761-44448	Sequence 44448, App
21	7	1.7	65	US-09-764-877-1958	Sequence 1958, App
22	7	1.7	117	US-10-156-761-13107	Sequence 13107, A
23	7	1.7	126	US-10-007-267-2	Sequence 2, Appli
24	7	1.7	132	US-09-956-423-4	Sequence 4, Appli
25	7	1.7	132	US-09-867-550-150	Sequence 150, App
26	7	1.7	153	US-09-925-300-1688	Sequence 1688, App
27	7	1.7	175	US-10-156-761-10928	Sequence 10928, A
28	7	1.7	184	US-10-156-761-10784	Sequence 10784, A
29	7	1.7	203	US-09-764-853-558	Sequence 558, App
30	7	1.7	236	US-09-738-626-6552	Sequence 6552, App
31	7	1.7	248	US-10-156-761-13873	Sequence 13873, A
32	7	1.7	253	US-10-317-806-4	Sequence 4, Appli
33	7	1.7	262	US-10-156-761-14170	Sequence 14170, A
34	7	1.7	271	US-10-342-224-72	Sequence 72, Appli
35	7	1.7	275	US-09-815-242-10842	Sequence 10842, A
36	7	1.7	275	US-10-080-170-328	Sequence 328, App
37	7	1.7	288	US-10-156-761-14182	Sequence 14182, A
38	7	1.7	300	US-10-156-761-8083	Sequence 8083, App
39	7	1.7	301	US-10-156-761-13108	Sequence 13108, App
40	7	1.7	302	US-09-533-029-28	Sequence 28, Appli
41	7	1.7	302	US-09-934-455-146	Sequence 146, App
42	7	1.7	302	US-10-295-403-86	Sequence 86, Appli
43	7	1.7	302	US-10-278-536-56	Sequence 56, Appli
44	7	1.7	304	US-10-156-761-14293	Sequence 14293, A
45	7	1.7	305	US-10-156-761-8219	Sequence 8219, App
46	7	1.7	308	US-09-738-626-4269	Sequence 4269, App
47	7	1.7	320	US-09-738-626-5385	Sequence 5385, App
48	7	1.7	339	US-09-802-371-2	Sequence 2, Appli
49	7	1.7	351	US-10-156-761-17774	Sequence 7774, App
50	7	1.7	361	US-10-114-893-88	Sequence 88, Appli
51	7	1.7	373	US-10-186-886-42	Sequence 42, Appli
52	7	1.7	373	US-10-186-886-43	Sequence 43, Appli
53	7	1.7	383	US-10-156-761-9603	Sequence 9603, App
54	7	1.7	396	US-09-712-363-231	Sequence 231, App
55	7	1.7	397	US-10-156-761-8560	Sequence 8560, App
56	7	1.7	434	US-10-102-806-506	Sequence 506, App
57	7	1.7	471	US-10-156-761-8883	Sequence 8883, App
58	7	1.7	471	US-10-156-761-8883	Sequence 8883, App
59	7	1.7	482	US-09-746-783-66	Sequence 66, Appli
60	7	1.7	487	US-10-137-870-528	Sequence 528, App
61	7	1.7	487	US-10-140-018-528	Sequence 528, App
62	7	1.7	487	US-10-140-021-528	Sequence 528, App
63	7	1.7	487	US-10-140-274-528	Sequence 528, App
64	7	1.7	487	US-10-140-471-528	Sequence 528, App
65	7	1.7	487	US-10-140-807-528	Sequence 528, App
66	7	1.7	487	US-10-140-922-528	Sequence 528, App
67	7	1.7	487	US-10-140-924-528	Sequence 528, App
68	7	1.7	487	US-10-140-926-528	Sequence 528, App
69	7	1.7	487	US-10-141-698-528	Sequence 528, App
70	7	1.7	487	US-10-141-702-528	Sequence 528, App
71	7	1.7	487	US-10-141-704-528	Sequence 528, App
72	7	1.7	487	US-10-142-421-528	Sequence 528, App
73	7	1.7	487	US-10-142-432-528	Sequence 528, App
74	7	1.7	487	US-10-142-767-528	Sequence 528, App
75	7	1.7	487	US-10-143-033-528	Sequence 528, App
76	7	1.7	487	US-10-143-994-528	Sequence 528, App
77	7	1.7	487	US-10-145-631-528	Sequence 528, App
78	7	1.7	487	US-10-145-631-528	Sequence 528, App
79	7	1.7	487	US-10-145-633-528	Sequence 528, App
80	7	1.7	487	US-10-145-746-528	Sequence 528, App
81	7	1.7	487	US-10-145-748-528	Sequence 528, App
82	7	1.7	487	US-10-145-823-528	Sequence 528, App
83	7	1.7	487	US-10-145-826-528	Sequence 528, App
84	7	1.7	487	US-10-145-870-528	Sequence 528, App
85	7	1.7	487	US-10-145-876-528	Sequence 528, App
86	7	1.7	487	US-10-145-895-528	Sequence 528, App
87	7	1.7	487	US-10-146-724-528	Sequence 528, App
88	7	1.7	487	US-10-146-725-528	Sequence 528, App

89	7	1.7	487	12	US-10-146-795-528	Sequence 528, App	162	7	1.7	487	15	US-10-033-435-18	Sequence 18, App1
90	7	1.7	487	12	US-10-147-495-528	Sequence 528, App	163	7	1.7	487	15	US-10-175-746-528	Sequence 528, App
91	7	1.7	487	12	US-10-147-501-528	Sequence 528, App	164	7	1.7	487	15	US-10-176-918-528	Sequence 528, App
92	7	1.7	487	12	US-10-147-504-528	Sequence 528, App	165	7	1.7	487	15	US-10-176-921-528	Sequence 528, App
93	7	1.7	487	12	US-10-147-506-528	Sequence 528, App	166	7	1.7	487	15	US-10-227-884-204	Sequence 204, App
94	7	1.7	487	12	US-10-147-509-528	Sequence 528, App	167	7	1.7	487	15	US-10-032-990-18	Sequence 18, App1
95	7	1.7	487	12	US-10-147-510-528	Sequence 528, App	168	7	1.7	487	15	US-10-137-865-528	Sequence 528, App
96	7	1.7	487	12	US-10-147-511-528	Sequence 528, App	169	7	1.7	487	15	US-10-140-474-528	Sequence 528, App
97	7	1.7	487	12	US-10-147-529-528	Sequence 528, App	170	7	1.7	487	15	US-10-142-431-528	Sequence 528, App
98	7	1.7	487	12	US-10-152-397-528	Sequence 528, App	171	7	1.7	487	15	US-10-143-114-528	Sequence 528, App
99	7	1.7	487	12	US-10-153-386-528	Sequence 528, App	172	7	1.7	487	15	US-10-230-163-204	Sequence 204, App
100	7	1.7	487	12	US-10-158-783-528	Sequence 528, App	173	7	1.7	487	15	US-10-140-002-528	Sequence 528, App
101	7	1.7	487	12	US-10-158-786-528	Sequence 528, App	174	7	1.7	487	15	US-10-230-338-204	Sequence 204, App
102	7	1.7	487	12	US-10-140-019-528	Sequence 528, App	175	7	1.7	487	15	US-10-142-419-528	Sequence 528, App
103	7	1.7	487	12	US-10-140-022-528	Sequence 528, App	176	7	1.7	487	15	US-10-218-631-204	Sequence 204, App
104	7	1.7	487	12	US-10-140-861-528	Sequence 528, App	177	7	1.7	487	15	US-10-133-262-528	Sequence 528, App
105	7	1.7	487	12	US-10-140-862-528	Sequence 528, App	178	7	1.7	487	15	US-10-142-423-528	Sequence 528, App
106	7	1.7	487	12	US-10-141-697-528	Sequence 528, App	179	7	1.7	487	15	US-10-230-414-204	Sequence 204, App
107	7	1.7	487	12	US-10-141-700-528	Sequence 528, App	180	7	1.7	487	15	US-10-032-996-18	Sequence 18, App1
108	7	1.7	487	12	US-10-141-705-528	Sequence 528, App	181	7	1.7	487	15	US-10-121-050-528	Sequence 528, App
109	7	1.7	487	12	US-10-141-753-528	Sequence 528, App	182	7	1.7	487	15	US-10-141-755-528	Sequence 528, App
110	7	1.7	487	12	US-10-141-758-528	Sequence 528, App	183	7	1.7	487	15	US-10-143-032-528	Sequence 528, App
111	7	1.7	487	12	US-10-142-418-528	Sequence 528, App	184	7	1.7	487	15	US-10-123-108-528	Sequence 528, App
112	7	1.7	487	12	US-10-142-420-528	Sequence 528, App	185	7	1.7	487	15	US-10-123-236-528	Sequence 528, App
113	7	1.7	487	12	US-10-142-422-528	Sequence 528, App	186	7	1.7	487	15	US-10-123-261-528	Sequence 528, App
114	7	1.7	487	12	US-10-142-427-528	Sequence 528, App	187	7	1.7	487	15	US-10-140-921-528	Sequence 528, App
115	7	1.7	487	12	US-10-143-760-528	Sequence 528, App	188	7	1.7	487	15	US-10-140-928-528	Sequence 528, App
116	7	1.7	487	12	US-10-145-821-528	Sequence 528, App	189	7	1.7	487	15	US-10-216-159A-204	Sequence 204, App
117	7	1.7	487	12	US-10-152-531-528	Sequence 528, App	190	7	1.7	487	15	US-10-121-045-528	Sequence 528, App
118	7	1.7	487	12	US-10-216-163-204	Sequence 204, App	191	7	1.7	487	15	US-10-123-282-528	Sequence 528, App
119	7	1.7	487	12	US-10-127-840A-528	Sequence 528, App	192	7	1.7	487	15	US-10-123-903-528	Sequence 528, App
120	7	1.7	487	12	US-10-143-761-528	Sequence 528, App	193	7	1.7	487	15	US-10-124-819-528	Sequence 528, App
121	7	1.7	487	12	US-10-143-763-528	Sequence 528, App	194	7	1.7	487	15	US-10-124-822-528	Sequence 528, App
122	7	1.7	487	12	US-10-142-763-528	Sequence 528, App	195	7	1.7	487	15	US-10-140-925-528	Sequence 528, App
123	7	1.7	487	12	US-10-142-765-528	Sequence 528, App	196	7	1.7	487	15	US-10-160-498-528	Sequence 528, App
124	7	1.7	487	12	US-10-142-887-528	Sequence 528, App	197	7	1.7	487	15	US-10-218-849-204	Sequence 204, App
125	7	1.7	487	12	US-10-142-888-528	Sequence 528, App	198	7	1.7	487	15	US-10-227-883-204	Sequence 204, App
126	7	1.7	487	12	US-10-143-034-528	Sequence 528, App	199	7	1.7	487	15	US-10-227-883-204	Sequence 204, App1
127	7	1.7	487	12	US-10-143-116-528	Sequence 528, App	200	7	1.7	487	15	US-10-033-396-18	Sequence 18, App1
128	7	1.7	487	12	US-10-144-957-528	Sequence 528, App	201	7	1.7	487	15	US-10-124-824-528	Sequence 528, App
129	7	1.7	487	12	US-10-144-992-528	Sequence 528, App	202	7	1.7	487	15	US-10-127-825A-528	Sequence 528, App
130	7	1.7	487	12	US-10-145-015-528	Sequence 528, App	203	7	1.7	487	15	US-10-127-829A-528	Sequence 528, App
131	7	1.7	487	12	US-10-145-090-528	Sequence 528, App	204	7	1.7	487	15	US-10-127-835A-528	Sequence 528, App
132	7	1.7	487	12	US-10-145-091-528	Sequence 528, App	205	7	1.7	487	15	US-10-131-823A-528	Sequence 528, App
133	7	1.7	487	12	US-10-145-629-528	Sequence 528, App	206	7	1.7	487	15	US-10-137-839A-528	Sequence 528, App
134	7	1.7	487	12	US-10-145-630-528	Sequence 528, App	207	7	1.7	487	15	US-10-128-693A-528	Sequence 528, App
135	7	1.7	487	12	US-10-145-747-528	Sequence 528, App	208	7	1.7	487	15	US-10-131-813A-528	Sequence 528, App
136	7	1.7	487	12	US-10-145-754-528	Sequence 528, App	209	7	1.7	487	15	US-10-131-818A-528	Sequence 528, App
137	7	1.7	487	12	US-10-145-754-528	Sequence 528, App	210	7	1.7	487	15	US-10-131-823A-528	Sequence 528, App
138	7	1.7	487	12	US-10-145-755-528	Sequence 528, App	211	7	1.7	487	15	US-10-131-824A-528	Sequence 528, App
139	7	1.7	487	12	US-10-145-818-528	Sequence 528, App	212	7	1.7	487	15	US-10-131-830A-528	Sequence 528, App
140	7	1.7	487	12	US-10-145-820-528	Sequence 528, App	213	7	1.7	487	15	US-10-137-872A-528	Sequence 528, App
141	7	1.7	487	12	US-10-145-872-528	Sequence 528, App	214	7	1.7	487	15	US-10-147-500-528	Sequence 528, App
142	7	1.7	487	12	US-10-145-873-528	Sequence 528, App	215	7	1.7	487	15	US-10-147-502-528	Sequence 528, App
143	7	1.7	487	12	US-10-147-481-528	Sequence 528, App	216	7	1.7	487	15	US-10-147-515-528	Sequence 528, App
144	7	1.7	487	12	US-10-147-482-528	Sequence 528, App	217	7	1.7	487	15	US-10-147-517-528	Sequence 528, App
145	7	1.7	487	12	US-10-147-503-528	Sequence 528, App	218	7	1.7	487	15	US-10-147-517-528	Sequence 528, App
146	7	1.7	487	12	US-10-147-522-528	Sequence 528, App	219	7	1.7	487	15	US-10-147-526-528	Sequence 528, App
147	7	1.7	487	12	US-10-152-401-528	Sequence 528, App	220	7	1.7	487	15	US-10-147-527-528	Sequence 528, App
148	7	1.7	487	12	US-10-157-783-528	Sequence 528, App	221	7	1.7	487	15	US-10-121-041-528	Sequence 528, App
149	7	1.7	487	12	US-10-158-463-528	Sequence 528, App	222	7	1.7	487	15	US-10-121-043-528	Sequence 528, App
150	7	1.7	487	12	US-10-158-792-528	Sequence 528, App	223	7	1.7	487	15	US-10-121-047-528	Sequence 528, App
151	7	1.7	487	13	US-10-033-246-18	Sequence 18, App1	224	7	1.7	487	15	US-10-123-215-528	Sequence 528, App
152	7	1.7	487	13	US-10-033-301-18	Sequence 18, App1	225	7	1.7	487	15	US-10-123-902-528	Sequence 528, App
153	7	1.7	487	13	US-10-033-326-18	Sequence 18, App1	226	7	1.7	487	15	US-10-123-908-528	Sequence 528, App
154	7	1.7	487	14	US-10-033-245-18	Sequence 18, App1	227	7	1.7	487	15	US-10-123-909-528	Sequence 528, App
155	7	1.7	487	14	US-10-033-223-18	Sequence 18, App1	228	7	1.7	487	15	US-10-123-910-528	Sequence 528, App
156	7	1.7	487	14	US-10-033-167-18	Sequence 18, App1	229	7	1.7	487	15	US-10-124-813-528	Sequence 528, App
157	7	1.7	487	14	US-10-033-244-18	Sequence 18, App1	230	7	1.7	487	15	US-10-124-817-528	Sequence 528, App
158	7	1.7	487	15	US-10-028-072-528	Sequence 528, App	231	7	1.7	487	15	US-10-125-922-528	Sequence 528, App
159	7	1.7	487	15	US-10-121-049-528	Sequence 528, App	232	7	1.7	487	15	US-10-125-924-528	Sequence 528, App
160	7	1.7	487	15	US-10-123-904-528	Sequence 528, App	233	7	1.7	487	15	US-10-140-860-528	Sequence 528, App
161	7	1.7	487	15	US-10-140-470-528	Sequence 528, App	234	7	1.7	487	15	US-10-142-417-528	Sequence 528, App

235	7	1.7	487	15	US-10-147-519-528	Sequence 528, App	308	7	1.7	487	15.	US-10-131-817A-528	Sequence 528, App
236	7	1.7	487	15	US-10-157-782-528	Sequence 528, App	309	7	1.7	487	15	US-10-131-821A-528	Sequence 528, App
237	7	1.7	487	15	US-10-152-395-528	Sequence 528, App	310	7	1.7	487	15	US-10-131-822A-528	Sequence 528, App
238	7	1.7	487	15	US-10-219-076-204	Sequence 204, App	311	7	1.7	487	15	US-10-131-828A-528	Sequence 528, App
239	7	1.7	487	15	US-10-230-434-204	Sequence 204, App	312	7	1.7	487	15	US-10-131-835A-528	Sequence 528, App
240	7	1.7	487	15	US-10-125-926A-528	Sequence 528, App	313	7	1.7	487	15	US-10-137-864A-528	Sequence 528, App
241	7	1.7	487	15	US-10-125-930A-528	Sequence 528, App	314	7	1.7	487	15	US-10-137-869A-528	Sequence 528, App
242	7	1.7	487	15	US-10-127-831A-528	Sequence 528, App	315	7	1.7	487	15	US-10-147-523-528	Sequence 528, App
243	7	1.7	487	15	US-10-127-837A-528	Sequence 528, App	316	7	1.7	487	15	US-10-158-785-528	Sequence 528, App
244	7	1.7	487	15	US-10-127-838B-528	Sequence 528, App	317	7	1.7	487	15	US-10-121-051-528	Sequence 528, App
245	7	1.7	487	15	US-10-127-842A-528	Sequence 528, App	318	7	1.7	487	15	US-10-216-165-204	Sequence 204, App
246	7	1.7	487	15	US-10-127-843A-528	Sequence 528, App	319	7	1.7	487	15	US-10-218-956-204	Sequence 204, App
247	7	1.7	487	15	US-10-127-845A-528	Sequence 528, App	320	7	1.7	487	15	US-10-219-466-204	Sequence 204, App
248	7	1.7	487	15	US-10-127-846A-528	Sequence 528, App	321	7	1.7	487	15	US-10-219-476-204	Sequence 204, App
249	7	1.7	487	15	US-10-127-848A-528	Sequence 528, App	322	7	1.7	487	15	US-10-219-536-204	Sequence 204, App
250	7	1.7	487	15	US-10-127-849A-528	Sequence 528, App	323	7	1.7	487	15	US-10-219-547A-204	Sequence 204, App
251	7	1.7	487	15	US-10-127-850A-528	Sequence 528, App	324	7	1.7	487	15	US-10-233-205-204	Sequence 204, App
252	7	1.7	487	15	US-10-127-851A-528	Sequence 528, App	325	7	1.7	487	15	US-10-121-042-528	Sequence 528, App
253	7	1.7	487	15	US-10-128-684A-528	Sequence 528, App	326	7	1.7	487	15	US-10-219-072-204	Sequence 204, App
254	7	1.7	487	15	US-10-128-686A-528	Sequence 528, App	327	7	1.7	487	15	US-10-219-474-204	Sequence 204, App
255	7	1.7	487	15	US-10-128-690A-528	Sequence 528, App	328	7	1.7	487	15	US-10-219-524-204	Sequence 204, App
256	7	1.7	487	15	US-10-128-691A-528	Sequence 528, App	329	7	1.7	487	15	US-10-219-528-204	Sequence 204, App
257	7	1.7	487	15	US-10-131-819A-528	Sequence 528, App	330	7	1.7	487	15	US-10-227-880-204	Sequence 204, App
258	7	1.7	487	15	US-10-131-829A-528	Sequence 528, App	331	7	1.7	487	15	US-10-227-881-204	Sequence 204, App
259	7	1.7	487	15	US-10-131-836A-528	Sequence 528, App	332	7	1.7	487	15	US-10-227-882-204	Sequence 204, App
260	7	1.7	487	15	US-10-146-729-528	Sequence 528, App	333	7	1.7	487	15	US-10-230-434-204	Sequence 204, App
261	7	1.7	487	15	US-10-146-791-528	Sequence 528, App	334	7	1.7	487	15	US-10-232-923-204	Sequence 204, App
262	7	1.7	487	15	US-10-147-484-528	Sequence 528, App	335	7	1.7	487	15	US-10-232-925-204	Sequence 204, App
263	7	1.7	487	15	US-10-147-508-528	Sequence 528, App	336	7	1.7	487	15	US-10-232-927-204	Sequence 204, App
264	7	1.7	487	15	US-10-147-512-528	Sequence 528, App	337	7	1.7	487	15	US-10-232-929-204	Sequence 204, App
265	7	1.7	487	15	US-10-175-735-528	Sequence 528, App	338	7	1.7	487	15	US-10-232-234-204	Sequence 204, App
266	7	1.7	487	15	US-10-121-040-528	Sequence 528, App	339	7	1.7	487	15	US-10-219-060-204	Sequence 204, App
267	7	1.7	487	15	US-10-121-056-528	Sequence 528, App	340	7	1.7	487	15	US-10-123-912-528	Sequence 528, App
268	7	1.7	487	15	US-10-121-061-528	Sequence 528, App	341	7	1.7	487	15	US-10-223-085-250	Sequence 250, App
269	7	1.7	487	15	US-10-123-235-528	Sequence 528, App	342	7	1.7	487	15	US-10-216-160-204	Sequence 204, App
270	7	1.7	487	15	US-10-124-818-528	Sequence 528, App	343	7	1.7	487	15	US-10-216-162-204	Sequence 204, App
271	7	1.7	487	15	US-10-137-868-528	Sequence 528, App	344	7	1.7	487	15	US-10-216-164-204	Sequence 204, App
272	7	1.7	487	15	US-10-147-492-528	Sequence 528, App	345	7	1.7	487	15	US-10-216-167-204	Sequence 204, App
273	7	1.7	487	15	US-10-158-782-528	Sequence 528, App	346	7	1.7	487	15.	US-10-216-168-204	Sequence 204, App
274	7	1.7	487	15	US-10-123-905-528	Sequence 528, App	347	7	1.7	487	15	US-10-219-065-204	Sequence 204, App
275	7	1.7	487	15	US-10-123-907-528	Sequence 528, App	348	7	1.7	487	15	US-10-219-071-204	Sequence 204, App
276	7	1.7	487	15	US-10-124-815-528	Sequence 528, App	349	7	1.7	487	15	US-10-219-074-204	Sequence 204, App
277	7	1.7	487	15	US-10-125-921A-528	Sequence 528, App	350	7	1.7	487	15	US-10-219-077-204	Sequence 204, App
278	7	1.7	487	15	US-10-125-928A-528	Sequence 528, App	351	7	1.7	487	15	US-10-219-465-204	Sequence 204, App
279	7	1.7	487	15	US-10-127-821A-528	Sequence 528, App	352	7	1.7	487	15	US-10-219-467-204	Sequence 204, App
280	7	1.7	487	15	US-10-127-822A-528	Sequence 528, App	353	7	1.7	487	15	US-10-219-468-204	Sequence 204, App
281	7	1.7	487	15	US-10-127-824A-528	Sequence 528, App	354	7	1.7	487	15	US-10-219-471-204	Sequence 204, App
282	7	1.7	487	15	US-10-127-826A-528	Sequence 528, App	355	7	1.7	487	15	US-10-219-473-204	Sequence 204, App
283	7	1.7	487	15	US-10-127-827A-528	Sequence 528, App	356	7	1.7	487	15	US-10-219-476-204	Sequence 204, App
284	7	1.7	487	15	US-10-127-828A-528	Sequence 528, App	357	7	1.7	487	15	US-10-219-487-204	Sequence 204, App
285	7	1.7	487	15	US-10-127-830A-528	Sequence 528, App	358	7	1.7	487	15	US-10-227-874-204	Sequence 204, App
286	7	1.7	487	15	US-10-127-832A-528	Sequence 528, App	359	7	1.7	487	15	US-10-227-876-204	Sequence 204, App
287	7	1.7	487	15	US-10-127-833A-528	Sequence 528, App	360	7	1.7	487	15	US-10-227-878-204	Sequence 204, App
288	7	1.7	487	15	US-10-127-834A-528	Sequence 528, App	361	7	1.7	487	15	US-10-227-879-204	Sequence 204, App
289	7	1.7	487	15	US-10-127-836A-528	Sequence 528, App	362	7	1.7	487	15	US-10-229-974-204	Sequence 204, App
290	7	1.7	487	15	US-10-127-841A-528	Sequence 528, App	363	7	1.7	487	15	US-10-230-024-204	Sequence 204, App
291	7	1.7	487	15	US-10-127-844A-528	Sequence 528, App	364	7	1.7	487	15	US-10-230-183-204	Sequence 204, App
292	7	1.7	487	15	US-10-128-687A-528	Sequence 528, App	365	7	1.7	487	15	US-10-230-234-204	Sequence 204, App
293	7	1.7	487	15	US-10-128-688A-528	Sequence 528, App	366	7	1.7	487	15	US-10-230-306-204	Sequence 204, App
294	7	1.7	487	15	US-10-128-689A-528	Sequence 528, App	367	7	1.7	487	15	US-10-230-426-204	Sequence 204, App
295	7	1.7	487	15	US-10-128-694A-528	Sequence 528, App	368	7	1.7	487	15	US-10-230-427-204	Sequence 204, App
296	7	1.7	487	15	US-10-131-825A-528	Sequence 528, App	369	7	1.7	487	15	US-10-230-433-204	Sequence 204, App
297	7	1.7	487	15	US-10-230-417-528	Sequence 528, App	370	7	1.7	487	15	US-10-230-435-204	Sequence 204, App
298	7	1.7	487	15	US-10-219-003-204	Sequence 204, App	371	7	1.7	487	15	US-10-230-438-204	Sequence 204, App
299	7	1.7	487	15	US-10-219-075-204	Sequence 204, App	372	7	1.7	487	15	US-10-233-222-204	Sequence 204, App
300	7	1.7	487	15	US-10-219-464-204	Sequence 204, App	373	7	1.7	487	15	US-10-199-007-528	Sequence 528, App
301	7	1.7	487	15	US-10-219-466-204	Sequence 204, App	374	7	1.7	487	15	US-10-199-034-528	Sequence 528, App
302	7	1.7	487	15	US-10-219-479-204	Sequence 204, App	375	7	1.7	487	15	US-10-223-088-250	Sequence 250, App
303	7	1.7	487	15	US-10-219-481-204	Sequence 204, App	376	7	1.7	487	15	US-10-223-088-250	Sequence 250, App
304	7	1.7	487	15	US-10-230-260-204	Sequence 204, App	377	7	1.7	487	15	US-10-223-090-250	Sequence 250, App
305	7	1.7	487	15	US-10-232-231-204	Sequence 204, App	378	7	1.7	487	15	US-10-219-070-204	Sequence 204, App
306	7	1.7	487	15	US-10-232-233-204	Sequence 204, App	379	7	1.7	487	15	US-10-219-472-204	Sequence 204, App
307	7	1.7	487	15	US-10-131-815A-528	Sequence 528, App	380	7	1.7	487	15	US-10-219-527-204	Sequence 204, App

Tue Sep 9 08:55:30 2003

527	6	1.5	70	10	US-09-289-346A-14	Sequence 14, Appl	600	6	1.5	142	15	US-10-156-761-7740	Sequence 7740, Ap
528	6	1.5	70	10	US-09-796-692-2365	Sequence 2365, Ap	601	6	1.5	142	15	US-10-156-761-10346	Sequence 10346, A
529	6	1.5	70	14	US-10-001-835-159	Sequence 159, App	602	6	1.5	142	15	US-10-156-761-12014	Sequence 12014, A
530	6	1.5	70	15	US-10-040-862-2365	Sequence 2365, App	603	6	1.5	143	10	US-09-738-626-5904	Sequence 5904, Ap
531	6	1.5	71	11	US-09-764-891-4600	Sequence 4600, Ap	604	6	1.5	143	15	US-10-156-761-9839	Sequence 9839, Ap
532	6	1.5	72	15	US-10-106-698-6127	Sequence 6127, Ap	605	6	1.5	144	11	US-09-764-881-133	Sequence 133, App
533	6	1.5	72	15	US-09-738-626-5102	Sequence 5102, Ap	606	6	1.5	147	10	US-09-764-864-1025	Sequence 1025, Ap
534	6	1.5	73	10	US-10-083-357-944	Sequence 944, App	607	6	1.5	149	9	US-09-940-497-2	Sequence 2, Appl1
535	6	1.5	74	11	US-09-974-879-244	Sequence 244, App	608	6	1.5	149	11	US-09-764-891-5349	Sequence 5349, Ap
536	6	1.5	75	11	US-09-305-736-245	Sequence 245, App	609	6	1.5	149	14	US-10-004-717-3	Sequence 23, Appl
537	6	1.5	77	11	US-09-918-036-3	Sequence 3, Appl1	610	6	1.5	150	10	US-09-731-872-363	Sequence 363, App
538	6	1.5	78	9	US-09-864-761-37650	Sequence 37650, A	611	6	1.5	150	10	US-09-738-626-4598	Sequence 4598, Ap
539	6	1.5	79	9	US-09-864-761-47232	Sequence 47232, A	612	6	1.5	150	12	US-09-876-997-363	Sequence 363, App
540	6	1.5	80	11	US-09-776-724A-139	Sequence 139, App	613	6	1.5	152	10	US-09-738-626-5856	Sequence 5856, Ap
541	6	1.5	80	11	US-09-948-820-72	Sequence 72, Appl	614	6	1.5	153	12	US-09-738-626-5856	Sequence 5856, Ap
542	6	1.5	86	15	US-10-011-585A-154	Sequence 154, App	615	6	1.5	156	12	US-10-017-161-1854	Sequence 1854, Ap
543	6	1.5	87	10	US-09-738-626-3513	Sequence 3513, Ap	616	6	1.5	157	12	US-10-017-161-1906	Sequence 1906, Ap
544	6	1.5	87	11	US-09-938-806A-5	Sequence 5, Appl1	617	6	1.5	157	15	US-10-156-761-10551	Sequence 10551, A
545	6	1.5	87	15	US-10-072-783-6	Sequence 6, Appl1	618	6	1.5	159	15	US-10-156-761-12728	Sequence 12728, A
546	6	1.5	88	11	US-09-764-891-3010	Sequence 3010, Ap	619	6	1.5	161	12	US-10-156-761-13807	Sequence 13807, A
547	6	1.5	90	10	US-09-818-247-17	Sequence 17, Appl	620	6	1.5	161	12	US-10-238-075-1254	Sequence 1254, Ap
548	6	1.5	90	11	US-09-818-247-17	Sequence 25, App	621	6	1.5	161	15	US-10-156-761-11117	Sequence 11117, A
549	6	1.5	90	14	US-10-001-873-34	Sequence 34, Appl	622	6	1.5	162	15	US-09-738-626-4267	Sequence 4267, A
550	6	1.5	94	10	US-09-925-300-1157	Sequence 1157, App	623	6	1.5	163	12	US-10-017-161-1894	Sequence 1894, Ap
551	6	1.5	94	11	US-09-764-872-412	Sequence 412, App	624	6	1.5	164	15	US-10-156-761-13229	Sequence 13229, A
552	6	1.5	95	9	US-09-925-301-892	Sequence 892, Appl	625	6	1.5	167	12	US-10-017-161-1154	Sequence 1154, Ap
553	6	1.5	95	10	US-09-818-247-20	Sequence 20, Appl	626	6	1.5	168	15	US-10-156-761-8638	Sequence 8638, Ap
554	6	1.5	96	11	US-09-764-891-1006	Sequence 3006, Ap	627	6	1.5	169	10	US-09-738-626-4159	Sequence 4159, Ap
555	6	1.5	96	14	US-10-093-892-7	Sequence 7, Appl1	628	6	1.5	170	10	US-09-839-894-36	Sequence 839, App
556	6	1.5	96	14	US-10-093-892-9	Sequence 9, Appl1	629	6	1.5	173	15	US-10-156-761-7665	Sequence 7665, Ap
557	6	1.5	96	14	US-10-093-892-9	Sequence 37100, A	630	6	1.5	175	9	US-09-815-242-11637	Sequence 11637, A
558	6	1.5	101	9	US-09-864-761-37100	Sequence 10, Appl	631	6	1.5	175	15	US-10-156-761-11876	Sequence 11876, A
559	6	1.5	102	12	US-09-969-748C-104	Sequence 20, Appl	632	6	1.5	175	15	US-10-156-761-9974	Sequence 9974, Ap
560	6	1.5	103	15	US-10-096-884-20	Sequence 5, Appl1	633	6	1.5	177	8	US-08-424-550B-49	Sequence 49, Appl
561	6	1.5	105	14	US-10-093-892-5	Sequence 84, Appl	634	6	1.5	177	8	US-09-963-347B-7	Sequence 7, Appl1
562	6	1.5	107	11	US-09-820-843A-84	Sequence 143, App	635	6	1.5	177	11	US-09-902-517-18	Sequence 18, Appl
563	6	1.5	108	15	US-10-078-030-143	Sequence 6169, Ap	636	6	1.5	177	11	US-09-902-517-18	Sequence 15, Appl
564	6	1.5	109	10	US-09-738-626-6169	Sequence 108, App	637	6	1.5	177	12	US-10-400-377-14	Sequence 75, Appl
565	6	1.5	112	11	US-09-948-820-108	Sequence 334, App	638	6	1.5	180	11	US-09-905-666A-75	Sequence 76, Appl
566	6	1.5	115	15	US-10-043-487-334	Sequence 8675, Ap	639	6	1.5	180	11	US-09-905-666A-76	Sequence 77, Appl
567	6	1.5	118	15	US-10-156-761-8675	Sequence 1172, Ap	640	6	1.5	180	11	US-09-905-666A-77	Sequence 78, Appl
568	6	1.5	119	10	US-09-764-868-1172	Sequence 14, Appl	641	6	1.5	180	11	US-09-905-666A-78	Sequence 79, Appl
569	6	1.5	121	10	US-09-775-932-14	Sequence 1641, App	642	6	1.5	180	11	US-09-905-666A-79	Sequence 80, Appl
570	6	1.5	121	10	US-09-764-877-1641	Sequence 1625, A	643	6	1.5	180	11	US-09-905-666A-80	Sequence 81, Appl
571	6	1.5	121	15	US-10-156-761-13625	Sequence 620, App	644	6	1.5	180	11	US-09-905-666A-81	Sequence 82, Appl
572	6	1.5	123	9	US-09-809-391-693	Sequence 6343, Ap	645	6	1.5	180	11	US-09-905-666A-82	Sequence 83, Appl
573	6	1.5	125	10	US-09-738-626-6343	Sequence 20, Appl	646	6	1.5	180	11	US-09-905-666A-83	Sequence 84, Appl
574	6	1.5	126	9	US-09-934-586A-20	Sequence 8612, Ap	647	6	1.5	180	11	US-09-905-666A-84	Sequence 85, Appl
575	6	1.5	126	15	US-10-156-761-8612	Sequence 12966, A	648	6	1.5	180	11	US-09-905-666A-85	Sequence 86, Appl
576	6	1.5	127	9	US-09-815-242-12966	Sequence 12, Appl	649	6	1.5	180	11	US-09-905-666A-86	Sequence 87, Appl
577	6	1.5	128	10	US-09-775-932-12	Sequence 538, App	650	6	1.5	180	11	US-09-905-666A-87	Sequence 88, Appl
578	6	1.5	130	9	US-09-764-853-538	Sequence 171, App	651	6	1.5	180	11	US-09-905-666A-88	Sequence 89, Appl
579	6	1.5	130	15	US-10-091-438-171	Sequence 10984, A	652	6	1.5	180	11	US-09-905-666A-89	Sequence 90, Appl
580	6	1.5	131	15	US-10-156-761-10984	Sequence 89, Appl	653	6	1.5	180	11	US-09-905-666A-90	Sequence 91, Appl
581	6	1.5	132	9	US-09-789-561-89	Sequence 474, App	654	6	1.5	180	11	US-09-905-666A-91	Sequence 92, Appl
582	6	1.5	134	10	US-09-731-872-474	Sequence 475, App	655	6	1.5	180	11	US-09-905-666A-92	Sequence 93, Appl
583	6	1.5	134	10	US-09-731-872-475	Sequence 474, App	656	6	1.5	180	11	US-09-905-666A-93	Sequence 94, Appl
584	6	1.5	134	10	US-09-738-626-5042	Sequence 474, App	657	6	1.5	180	11	US-09-905-666A-94	Sequence 95, Appl
585	6	1.5	134	12	US-09-876-997-474	Sequence 475, App	658	6	1.5	180	11	US-09-905-666A-95	Sequence 96, Appl
586	6	1.5	134	12	US-09-876-997-475	Sequence 123, App	659	6	1.5	180	11	US-09-905-666A-96	Sequence 97, Appl
587	6	1.5	136	10	US-09-738-973-123	Sequence 123, App	660	6	1.5	180	11	US-09-905-666A-97	Sequence 98, Appl
588	6	1.5	136	10	US-09-854-133-123	Sequence 123, App	661	6	1.5	180	11	US-09-905-666A-98	Sequence 99, Appl
589	6	1.5	136	15	US-10-144-649A-123	Sequence 12625, A	662	6	1.5	180	11	US-09-905-666A-99	Sequence 99, Appl
590	6	1.5	137	15	US-10-156-761-12625	Sequence 210, App	663	6	1.5	180	11	US-09-905-666A-100	Sequence 100, App
591	6	1.5	138	11	US-09-468-147-210	Sequence 210, App	664	6	1.5	180	11	US-09-905-666A-101	Sequence 101, App
592	6	1.5	138	11	US-09-468-147-214	Sequence 248, App	665	6	1.5	180	11	US-09-905-666A-102	Sequence 102, App
593	6	1.5	138	11	US-09-468-147-244	Sequence 248, App	666	6	1.5	180	11	US-09-905-666A-103	Sequence 103, App
594	6	1.5	138	11	US-09-468-147-248	Sequence 252, App	667	6	1.5	180	11	US-09-905-666A-104	Sequence 104, App
595	6	1.5	138	11	US-09-468-147-252	Sequence 142, App	668	6	1.5	180	11	US-09-905-666A-105	Sequence 105, App
596	6	1.5	138	11	US-09-468-147-252	Sequence 289, App	669	6	1.5	180	11	US-09-905-666A-106	Sequence 106, App
597	6	1.5	140	9	US-09-764-881-142	Sequence 66, Appl	670	6	1.5	180	11	US-09-905-666A-107	Sequence 108, App
598	6	1.5	140	12	US-09-890-688-68	Sequence 289, App	671	6	1.5	180	11	US-09-905-666A-108	Sequence 108, App
599	6	1.5	140	15	US-10-073-961-289	Sequence 289, App	672	6	1.5	181	11	US-09-848-616-155	Sequence 155, App

673	6	1.5	181	12	US-10-289-454-6	Sequence 6, Appli	746	6	1.5	204	15	US-10-180-557-40	Sequence 40, Appl
674	6	1.5	181	15	US-10-156-761-13007	Sequence 13007, A	747	6	1.5	204	15	US-10-173-700-40	Sequence 40, Appl
675	6	1.5	182	9	US-09-912-020-361	Sequence 361, App	748	6	1.5	204	15	US-10-174-572-40	Sequence 40, Appl
676	6	1.5	182	9	US-09-815-242-10472	Sequence 10472, A	749	6	1.5	204	15	US-10-174-579-40	Sequence 40, Appl
677	6	1.5	182	11	US-09-848-616-184	Sequence 184, App	750	6	1.5	204	15	US-10-174-582-40	Sequence 40, Appl
678	6	1.5	183	14	US-10-033-109-2	Sequence 2, Appli	751	6	1.5	204	15	US-10-174-588-40	Sequence 40, Appl
679	6	1.5	186	9	US-09-815-242-5112	Sequence 5112, Ap	752	6	1.5	204	15	US-10-175-739-40	Sequence 40, Appl
680	6	1.5	186	15	US-10-156-761-10927	Sequence 10927, A	753	6	1.5	204	15	US-10-175-740-40	Sequence 40, Appl
681	6	1.5	188	15	US-10-156-761-10515	Sequence 10515, A	754	6	1.5	204	15	US-10-175-743-40	Sequence 40, Appl
682	6	1.5	191	10	US-09-944-849-7	Sequence 7, Appli	755	6	1.5	204	15	US-10-176-488-40	Sequence 40, Appl
683	6	1.5	192	15	US-10-156-761-11593	Sequence 11593, A	756	6	1.5	204	15	US-10-176-492-40	Sequence 40, Appl
684	6	1.5	193	9	US-09-764-898-214	Sequence 214, App	757	6	1.5	204	15	US-10-176-747-40	Sequence 40, Appl
685	6	1.5	193	10	US-09-738-626-5496	Sequence 5496, Ap	758	6	1.5	204	15	US-10-176-750-40	Sequence 40, Appl
686	6	1.5	193	15	US-10-156-761-7869	Sequence 7869, App	759	6	1.5	204	15	US-10-176-885-40	Sequence 40, Appl
687	6	1.5	195	15	US-10-156-761-10076	Sequence 10076, A	760	6	1.5	204	15	US-10-176-887-40	Sequence 40, Appl
688	6	1.5	196	14	US-10-025-367-4	Sequence 4, Appli	761	6	1.5	204	15	US-10-176-892-40	Sequence 40, Appl
689	6	1.5	196	15	US-10-207-655-143	Sequence 143, App	762	6	1.5	204	15	US-10-176-893-40	Sequence 40, Appl
690	6	1.5	199	15	US-10-152-300-2	Sequence 2, Appli	763	6	1.5	204	15	US-10-176-898-40	Sequence 40, Appl
691	6	1.5	201	10	US-09-901-540-3	Sequence 3, Appli	764	6	1.5	204	15	US-10-176-899-40	Sequence 40, Appl
692	6	1.5	201	10	US-09-896-856-8	Sequence 8, Appli	765	6	1.5	204	15	US-10-173-695-40	Sequence 40, Appl
693	6	1.5	201	10	US-09-901-257-3	Sequence 3, Appli	766	6	1.5	204	15	US-10-173-697-40	Sequence 40, Appl
694	6	1.5	201	15	US-10-107-931-8	Sequence 8, Appli	767	6	1.5	204	15	US-10-173-705-40	Sequence 40, Appl
695	6	1.5	201	15	US-10-212-793-4	Sequence 4, Appli	768	6	1.5	204	15	US-10-174-576-40	Sequence 40, Appl
696	6	1.5	201	15	US-10-156-761-12064	Sequence 12064, A	769	6	1.5	204	15	US-10-174-585-40	Sequence 40, Appl
697	6	1.5	203	15	US-10-156-761-12234	Sequence 12234, A	770	6	1.5	204	15	US-10-174-586-40	Sequence 40, Appl
698	6	1.5	204	9	US-09-778-320-295	Sequence 295, App	771	6	1.5	204	15	US-10-175-747-40	Sequence 40, Appl
699	6	1.5	204	9	US-09-910-689-295	Sequence 295, App	772	6	1.5	204	15	US-10-176-481-40	Sequence 40, Appl
700	6	1.5	204	10	US-09-978-295A-36	Sequence 36, Appl	773	6	1.5	204	15	US-10-176-485-40	Sequence 40, Appl
701	6	1.5	204	10	US-09-978-697-36	Sequence 36, Appl	774	6	1.5	204	15	US-10-176-487-40	Sequence 40, Appl
702	6	1.5	204	10	US-09-978-192A-36	Sequence 36, Appl	775	6	1.5	204	15	US-10-176-493-40	Sequence 40, Appl
703	6	1.5	204	10	US-09-999-832A-36	Sequence 36, Appl	776	6	1.5	204	15	US-10-176-756-40	Sequence 40, Appl
704	6	1.5	204	11	US-09-978-189-36	Sequence 36, Appl	777	6	1.5	204	15	US-10-176-911-40	Sequence 40, Appl
705	6	1.5	204	11	US-09-978-608A-36	Sequence 36, Appl	778	6	1.5	204	15	US-10-176-919-40	Sequence 40, Appl
706	6	1.5	204	11	US-09-809-391-429	Sequence 429, App	779	6	1.5	204	15	US-10-176-925-40	Sequence 40, Appl
707	6	1.5	204	11	US-09-978-585A-36	Sequence 36, Appl	780	6	1.5	204	15	US-10-176-978-40	Sequence 40, Appl
708	6	1.5	204	11	US-09-978-191A-36	Sequence 36, Appl	781	6	1.5	204	15	US-10-179-10-40	Sequence 40, Appl
709	6	1.5	204	11	US-09-978-403A-36	Sequence 36, Appl	782	6	1.5	204	15	US-10-180-543-40	Sequence 40, Appl
710	6	1.5	204	11	US-09-978-423A-36	Sequence 36, Appl	783	6	1.5	204	15	US-10-180-544-40	Sequence 40, Appl
711	6	1.5	204	11	US-09-999-833A-36	Sequence 36, Appl	784	6	1.5	204	15	US-10-180-546-40	Sequence 40, Appl
712	6	1.5	204	11	US-09-981-915A-36	Sequence 36, Appl	785	6	1.5	204	15	US-10-180-547-40	Sequence 40, Appl
713	6	1.5	204	11	US-09-978-824-36	Sequence 36, Appl	786	6	1.5	204	15	US-10-180-549-40	Sequence 40, Appl
714	6	1.5	204	11	US-09-918-585A-36	Sequence 36, Appl	787	6	1.5	204	15	US-10-180-555-40	Sequence 40, Appl
715	6	1.5	204	11	US-09-978-423A-36	Sequence 36, Appl	788	6	1.5	204	15	US-10-180-559-40	Sequence 40, Appl
716	6	1.5	204	11	US-09-978-193A-36	Sequence 36, Appl	789	6	1.5	204	15	US-10-181-000-40	Sequence 40, Appl
717	6	1.5	204	11	US-09-999-830A-36	Sequence 36, Appl	790	6	1.5	204	15	US-10-183-010-40	Sequence 40, Appl
718	6	1.5	204	11	US-09-978-757A-36	Sequence 36, Appl	791	6	1.5	204	15	US-10-183-012-40	Sequence 40, Appl
719	6	1.5	204	11	US-09-978-187B-36	Sequence 36, Appl	792	6	1.5	204	15	US-10-184-614-40	Sequence 40, Appl
720	6	1.5	204	11	US-09-978-643A-36	Sequence 36, Appl	793	6	1.5	204	15	US-10-184-623-40	Sequence 40, Appl
721	6	1.5	204	12	US-09-978-375A-36	Sequence 36, Appl	794	6	1.5	204	15	US-10-184-635-40	Sequence 40, Appl
722	6	1.5	204	12	US-09-978-188A-36	Sequence 36, Appl	795	6	1.5	204	15	US-10-184-637-40	Sequence 40, Appl
723	6	1.5	204	12	US-09-978-298A-36	Sequence 36, Appl	796	6	1.5	204	15	US-10-184-646-40	Sequence 40, Appl
724	6	1.5	204	12	US-10-143-031A-36	Sequence 36, Appl	797	6	1.5	204	15	US-10-184-647-40	Sequence 40, Appl
725	6	1.5	204	12	US-10-002-967A-36	Sequence 36, Appl	798	6	1.5	204	15	US-10-184-652-40	Sequence 40, Appl
726	6	1.5	204	12	US-10-017-083A-36	Sequence 36, Appl	799	6	1.5	204	15	US-10-187-594-40	Sequence 40, Appl
727	6	1.5	204	12	US-10-143-030A-36	Sequence 36, Appl	800	6	1.5	204	15	US-10-187-596-40	Sequence 40, Appl
728	6	1.5	204	12	US-10-199-672-40	Sequence 40, Appl	801	6	1.5	204	15	US-10-187-745-40	Sequence 40, Appl
729	6	1.5	204	12	US-10-187-749-40	Sequence 40, Appl	802	6	1.5	204	15	US-10-187-885-40	Sequence 40, Appl
730	6	1.5	204	12	US-10-194-457-40	Sequence 40, Appl	803	6	1.5	204	15	US-10-187-886-40	Sequence 40, Appl
731	6	1.5	204	12	US-10-145-128A-36	Sequence 36, Appl	804	6	1.5	204	15	US-10-189-464-40	Sequence 40, Appl
732	6	1.5	204	12	US-10-184-642-40	Sequence 40, Appl	805	6	1.5	204	15	US-10-189-465-40	Sequence 40, Appl
733	6	1.5	204	12	US-10-196-747-40	Sequence 40, Appl	806	6	1.5	204	15	US-10-196-756-40	Sequence 40, Appl
734	6	1.5	204	14	US-10-052-586-40	Sequence 40, Appl	807	6	1.5	204	15	US-10-176-751-40	Sequence 40, Appl
735	6	1.5	204	14	US-10-010-742-285	Sequence 295, App	808	6	1.5	204	15	US-10-176-960-40	Sequence 40, Appl
736	6	1.5	204	15	US-10-174-590-40	Sequence 40, Appl	809	6	1.5	204	15	US-10-180-541-40	Sequence 40, Appl
737	6	1.5	204	15	US-10-176-758-40	Sequence 40, Appl	810	6	1.5	204	15	US-10-180-542-40	Sequence 40, Appl
738	6	1.5	204	15	US-10-175-737-40	Sequence 40, Appl	811	6	1.5	204	15	US-10-180-548-40	Sequence 40, Appl
739	6	1.5	204	15	US-10-173-706-40	Sequence 40, Appl	812	6	1.5	204	15	US-10-180-551-40	Sequence 40, Appl
740	6	1.5	204	15	US-10-175-738-40	Sequence 40, Appl	813	6	1.5	204	15	US-10-180-998-40	Sequence 40, Appl
741	6	1.5	204	15	US-10-175-752-40	Sequence 40, Appl	814	6	1.5	204	15	US-10-180-999-40	Sequence 40, Appl
742	6	1.5	204	15	US-10-176-482-40	Sequence 40, Appl	815	6	1.5	204	15	US-10-183-013-40	Sequence 40, Appl
743	6	1.5	204	15	US-10-176-757-40	Sequence 40, Appl	816	6	1.5	204	15	US-10-184-612-40	Sequence 40, Appl
744	6	1.5	204	15	US-10-176-913-40	Sequence 40, Appl	817	6	1.5	204	15	US-10-184-616-40	Sequence 40, Appl
745	6	1.5	204	15	US-10-180-552-40	Sequence 40, Appl	818	6	1.5	204	15	US-10-184-617-40	Sequence 40, Appl

819	6	1.5	204	15	US-10-184-622-40	Sequence 40, Appl	892	6	1.5	204	15	US-10-179-522-40	Sequence 40, Appl
820	6	1.5	204	15	US-10-184-628-40	Sequence 40, Appl	893	6	1.5	204	15	US-10-180-556-40	Sequence 40, Appl
821	6	1.5	204	15	US-10-184-629-40	Sequence 40, Appl	894	6	1.5	204	15	US-10-180-560-40	Sequence 40, Appl
822	6	1.5	204	15	US-10-184-630-40	Sequence 40, Appl	895	6	1.5	204	15	US-10-183-015-40	Sequence 40, Appl
823	6	1.5	204	15	US-10-184-631-40	Sequence 40, Appl	896	6	1.5	204	15	US-10-184-615-40	Sequence 40, Appl
824	6	1.5	204	15	US-10-184-632-40	Sequence 40, Appl	897	6	1.5	204	15	US-10-184-620-40	Sequence 40, Appl
825	6	1.5	204	15	US-10-184-636-40	Sequence 40, Appl	898	6	1.5	204	15	US-10-184-643-40	Sequence 40, Appl
826	6	1.5	204	15	US-10-184-640-40	Sequence 40, Appl	899	6	1.5	204	15	US-10-184-650-40	Sequence 40, Appl
827	6	1.5	204	15	US-10-184-650-40	Sequence 40, Appl	900	6	1.5	204	15	US-10-192-010-40	Sequence 40, Appl
828	6	1.5	204	15	US-10-184-651-40	Sequence 40, Appl	901	6	1.5	204	15	US-10-205-908-40	Sequence 40, Appl
829	6	1.5	204	15	US-10-187-588-40	Sequence 40, Appl	902	6	1.5	204	15	US-10-186-885-40	Sequence 40, Appl
830	6	1.5	204	15	US-10-187-597-40	Sequence 40, Appl	903	6	1.5	204	15	US-10-017-681A-6	Sequence 36, Appl
831	6	1.5	204	15	US-10-187-598-40	Sequence 40, Appl	904	6	1.5	204	15	US-10-184-619-40	Sequence 40, Appl
832	6	1.5	204	15	US-10-187-601-40	Sequence 40, Appl	905	6	1.5	204	15	US-10-187-599-40	Sequence 40, Appl
833	6	1.5	204	15	US-10-187-601-40	Sequence 40, Appl	906	6	1.5	204	15	US-10-187-750-40	Sequence 40, Appl
834	6	1.5	204	15	US-10-187-602-40	Sequence 40, Appl	907	6	1.5	204	15	US-10-188-780-40	Sequence 40, Appl
835	6	1.5	204	15	US-10-187-603-40	Sequence 40, Appl	908	6	1.5	204	15	US-10-192-015-40	Sequence 40, Appl
836	6	1.5	204	15	US-10-187-741-40	Sequence 40, Appl	909	6	1.5	204	15	US-10-194-394-40	Sequence 40, Appl
837	6	1.5	204	15	US-10-187-743-40	Sequence 40, Appl	910	6	1.5	204	15	US-10-194-425-40	Sequence 40, Appl
838	6	1.5	204	15	US-10-187-746-40	Sequence 40, Appl	911	6	1.5	204	15	US-10-194-485-40	Sequence 40, Appl
839	6	1.5	204	15	US-10-187-747-40	Sequence 40, Appl	912	6	1.5	204	15	US-10-195-885-40	Sequence 40, Appl
840	6	1.5	204	15	US-10-187-751-40	Sequence 40, Appl	913	6	1.5	204	15	US-10-195-889-40	Sequence 40, Appl
841	6	1.5	204	15	US-10-187-753-40	Sequence 40, Appl	914	6	1.5	204	15	US-10-196-748-40	Sequence 40, Appl
842	6	1.5	204	15	US-10-187-754-40	Sequence 40, Appl	915	6	1.5	204	15	US-10-196-750-40	Sequence 40, Appl
843	6	1.5	204	15	US-10-187-757-40	Sequence 40, Appl	916	6	1.5	204	15	US-10-197-659-40	Sequence 40, Appl
844	6	1.5	204	15	US-10-187-784-40	Sequence 40, Appl	917	6	1.5	204	15	US-10-197-700-40	Sequence 40, Appl
845	6	1.5	204	15	US-10-188-767-40	Sequence 40, Appl	918	6	1.5	204	15	US-10-197-705-40	Sequence 40, Appl
846	6	1.5	204	15	US-10-188-769-40	Sequence 40, Appl	919	6	1.5	204	15	US-10-197-708-40	Sequence 40, Appl
847	6	1.5	204	15	US-10-188-770-40	Sequence 40, Appl	920	6	1.5	204	15	US-10-198-764-40	Sequence 40, Appl
848	6	1.5	204	15	US-10-188-773-40	Sequence 40, Appl	921	6	1.5	204	15	US-10-198-765-40	Sequence 40, Appl
849	6	1.5	204	15	US-10-188-781-40	Sequence 40, Appl	922	6	1.5	204	15	US-10-198-769-40	Sequence 40, Appl
850	6	1.5	204	15	US-10-194-361-40	Sequence 40, Appl	923	6	1.5	204	15	US-10-198-768-40	Sequence 40, Appl
851	6	1.5	204	15	US-10-194-493-40	Sequence 40, Appl	924	6	1.5	204	15	US-10-199-305-40	Sequence 40, Appl
852	6	1.5	204	15	US-10-195-897-40	Sequence 40, Appl	925	6	1.5	204	15	US-10-199-306-40	Sequence 40, Appl
853	6	1.5	204	15	US-10-195-901-40	Sequence 40, Appl	926	6	1.5	204	15	US-10-199-310-40	Sequence 40, Appl
854	6	1.5	204	15	US-10-195-902-40	Sequence 40, Appl	927	6	1.5	204	15	US-10-199-311-40	Sequence 40, Appl
855	6	1.5	204	15	US-10-196-743-40	Sequence 40, Appl	928	6	1.5	204	15	US-10-199-314-40	Sequence 40, Appl
856	6	1.5	204	15	US-10-196-760-40	Sequence 40, Appl	929	6	1.5	204	15	US-10-199-317-40	Sequence 40, Appl
857	6	1.5	204	15	US-10-173-708-40	Sequence 40, Appl	930	6	1.5	204	15	US-10-199-665-40	Sequence 40, Appl
858	6	1.5	204	15	US-10-176-479-40	Sequence 40, Appl	931	6	1.5	204	15	US-10-199-666-40	Sequence 40, Appl
859	6	1.5	204	15	US-10-176-748-40	Sequence 40, Appl	932	6	1.5	204	15	US-10-199-669-40	Sequence 40, Appl
860	6	1.5	204	15	US-10-176-916-40	Sequence 40, Appl	933	6	1.5	204	15	US-10-201-534-40	Sequence 40, Appl
861	6	1.5	204	15	US-10-179-507-40	Sequence 40, Appl	934	6	1.5	204	15	US-10-201-770-40	Sequence 40, Appl
862	6	1.5	204	15	US-10-179-516-40	Sequence 40, Appl	935	6	1.5	204	15	US-10-201-855-40	Sequence 40, Appl
863	6	1.5	204	15	US-10-179-519-40	Sequence 40, Appl	936	6	1.5	204	15	US-10-201-856-40	Sequence 40, Appl
864	6	1.5	204	15	US-10-179-525-40	Sequence 40, Appl	937	6	1.5	204	15	US-10-202-469-40	Sequence 40, Appl
865	6	1.5	204	15	US-10-180-540-40	Sequence 40, Appl	938	6	1.5	204	15	US-10-202-470-40	Sequence 40, Appl
866	6	1.5	204	15	US-10-180-545-40	Sequence 40, Appl	939	6	1.5	204	15	US-10-202-476-40	Sequence 40, Appl
867	6	1.5	204	15	US-10-183-006-40	Sequence 40, Appl	940	6	1.5	204	15	US-10-202-934-40	Sequence 40, Appl
868	6	1.5	204	15	US-10-183-008-40	Sequence 40, Appl	941	6	1.5	204	15	US-10-202-935-40	Sequence 40, Appl
869	6	1.5	204	15	US-10-183-017-40	Sequence 40, Appl	942	6	1.5	204	15	US-10-202-936-40	Sequence 40, Appl
870	6	1.5	204	15	US-10-183-019-40	Sequence 40, Appl	943	6	1.5	204	15	US-10-202-939-40	Sequence 40, Appl
871	6	1.5	204	15	US-10-184-618-40	Sequence 40, Appl	944	6	1.5	204	15	US-10-205-504-40	Sequence 40, Appl
872	6	1.5	204	15	US-10-184-625-40	Sequence 40, Appl	945	6	1.5	204	15	US-10-205-509-40	Sequence 40, Appl
873	6	1.5	204	15	US-10-184-627-40	Sequence 40, Appl	946	6	1.5	204	15	US-10-205-895-40	Sequence 40, Appl
874	6	1.5	204	15	US-10-184-627-40	Sequence 40, Appl	947	6	1.5	204	15	US-10-205-899-40	Sequence 40, Appl
875	6	1.5	204	15	US-10-184-645-40	Sequence 40, Appl	948	6	1.5	204	15	US-10-205-900-40	Sequence 40, Appl
876	6	1.5	204	15	US-10-184-654-40	Sequence 40, Appl	949	6	1.5	204	15	US-10-205-909-40	Sequence 40, Appl
877	6	1.5	204	15	US-10-184-655-40	Sequence 40, Appl	950	6	1.5	204	15	US-10-195-880-40	Sequence 40, Appl
878	6	1.5	204	15	US-10-188-774-40	Sequence 40, Appl	951	6	1.5	204	15	US-10-183-002-40	Sequence 40, Appl
879	6	1.5	204	15	US-10-188-775-40	Sequence 40, Appl	952	6	1.5	204	15	US-10-184-621-40	Sequence 40, Appl
880	6	1.5	204	15	US-10-194-462-40	Sequence 40, Appl	953	6	1.5	204	15	US-10-184-638-40	Sequence 40, Appl
881	6	1.5	204	15	US-10-196-745-40	Sequence 40, Appl	954	6	1.5	204	15	US-10-187-752-40	Sequence 40, Appl
882	6	1.5	204	15	US-10-196-762-40	Sequence 40, Appl	955	6	1.5	204	15	US-10-187-887-40	Sequence 40, Appl
883	6	1.5	204	15	US-10-197-635-40	Sequence 40, Appl	956	6	1.5	204	15	US-10-194-461-40	Sequence 40, Appl
884	6	1.5	204	15	US-10-195-894-40	Sequence 40, Appl	957	6	1.5	204	15	US-10-195-892-40	Sequence 40, Appl
885	6	1.5	204	15	US-10-176-484-40	Sequence 40, Appl	958	6	1.5	204	15	US-10-196-751-40	Sequence 40, Appl
886	6	1.5	204	15	US-10-176-753-40	Sequence 40, Appl	959	6	1.5	204	15	US-10-197-694-40	Sequence 40, Appl
887	6	1.5	204	15	US-10-176-917-40	Sequence 40, Appl	960	6	1.5	204	15	US-10-197-697-40	Sequence 40, Appl
888	6	1.5	204	15	US-10-179-982-40	Sequence 40, Appl	961	6	1.5	204	15	US-10-197-707-40	Sequence 40, Appl
889	6	1.5	204	15	US-10-179-506-40	Sequence 40, Appl	962	6	1.5	204	15	US-10-199-303-40	Sequence 40, Appl
890	6	1.5	204	15	US-10-179-513-40	Sequence 40, Appl	963	6	1.5	204	15	US-10-199-318-40	Sequence 40, Appl
891	6	1.5	204	15	US-10-179-514-40	Sequence 40, Appl	964	6	1.5	204	15	US-10-199-458-40	Sequence 40, Appl

```
965 6 1.5 204 15 US-10-199-462-40 Sequence 40, Appl
966 6 1.5 204 15 US-10-201-324-40 Sequence 40, Appl
967 6 1.5 204 15 US-10-201-328-40 Sequence 40, Appl
968 6 1.5 204 15 US-10-201-527-40 Sequence 40, Appl
969 6 1.5 204 15 US-10-201-528-40 Sequence 40, Appl
970 6 1.5 204 15 US-10-201-529-40 Sequence 40, Appl
971 6 1.5 204 15 US-10-201-530-40 Sequence 40, Appl
972 6 1.5 204 15 US-10-202-408-40 Sequence 40, Appl
973 6 1.5 204 15 US-10-202-409-40 Sequence 40, Appl
974 6 1.5 204 15 US-10-202-411-40 Sequence 40, Appl
975 6 1.5 204 15 US-10-202-472-40 Sequence 40, Appl
976 6 1.5 204 15 US-10-205-502-40 Sequence 40, Appl
977 6 1.5 204 15 US-10-205-507-40 Sequence 40, Appl
978 6 1.5 204 15 US-10-205-511-40 Sequence 40, Appl
979 6 1.5 204 15 US-10-205-902-40 Sequence 40, Appl
980 6 1.5 204 15 US-10-205-907-40 Sequence 40, Appl
981 6 1.5 204 15 US-10-167-749-36 Sequence 36, Appl
982 6 1.5 204 15 US-10-194-456-40 Sequence 40, Appl
983 6 1.5 204 15 US-10-196-758-40 Sequence 40, Appl
984 6 1.5 204 15 US-10-198-770-40 Sequence 40, Appl
985 6 1.5 204 15 US-10-198-308-40 Sequence 40, Appl
986 6 1.5 204 15 US-10-200-617-40 Sequence 40, Appl
987 6 1.5 204 15 US-10-205-893-40 Sequence 40, Appl
988 6 1.5 204 15 US-10-205-897-40 Sequence 40, Appl
989 6 1.5 204 15 US-10-196-754-40 Sequence 40, Appl
990 6 1.5 204 15 US-10-176-847-8 Sequence 8, Appl
991 6 1.5 204 15 US-10-013-921A-36 Sequence 36, Appl
992 6 1.5 204 15 US-10-174-571-40 Sequence 40, Appl
993 6 1.5 204 15 US-10-176-746-40 Sequence 40, Appl
994 6 1.5 204 15 US-10-176-923-40 Sequence 40, Appl
995 6 1.5 204 15 US-10-183-011-40 Sequence 40, Appl
996 6 1.5 204 15 US-10-184-633-40 Sequence 40, Appl
997 6 1.5 204 15 US-10-184-639-40 Sequence 40, Appl
998 6 1.5 204 15 US-10-187-742-40 Sequence 40, Appl
999 6 1.5 204 15 US-10-187-748-40 Sequence 40, Appl
1000 6 1.5 204 15 US-10-206-907-40 Sequence 40, Appl
```

ALIGNMENTS

```
RESULT 1
US-10-012-542-398
; Sequence 398, Application US/10012542
; Publication No. US20030044851A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 94 Human Secreted Proteins
; FILE REFERENCE: P2029P1
; CURRENT APPLICATION NUMBER: US/10/012,542
; CURRENT FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/461,325
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-12-14
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,507
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,508
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,509
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,510
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,112
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,113
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22
; NUMBER OF SEQ ID NOS: 532
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 398
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-542-398
```

```
Query Match 2.4%; Score 10; DB 15; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 10 RAALAAVA 19
|||||
Db 104 RAALAAVA 113
```

```
RESULT 2
US-09-890-688-154
; Sequence 154, Application US/09890688
; Publication No. US20030144475A1
; GENERAL INFORMATION:
; APPLICANT: Seiichi KATO
; APPLICANT: Chikashi EGUCHI
; APPLICANT: Mihiro SAKETI
; TITLE OF INVENTION: Human Proteins and cDNAs thereof
; FILE REFERENCE: 2001-1102A/MMC/00653
; CURRENT APPLICATION NUMBER: US/09/890,688
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: JP 11-346863
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: JP 11-34684
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: JP 2000-31062
; PRIOR FILING DATE: 2000-02-08
; PRIOR APPLICATION NUMBER: JP 2000-34091
; PRIOR FILING DATE: 2000-02-10
; PRIOR APPLICATION NUMBER: JP 2000-34090
; PRIOR FILING DATE: 2000-02-10
; PRIOR APPLICATION NUMBER: JP 2000-35829
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: JP 2000-35899
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: JP 2000-71161
; PRIOR FILING DATE: 2000-03-14
; PRIOR APPLICATION NUMBER: JP 2000-160851
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 154
; LENGTH: 222
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-890-688-154

Query Match 2.4%; Score 10; DB 12; Length 222;
Best Local Similarity 100.0%; Pred. No. 0.28;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 10 RAALAAVA 19
|||||
Db 202 RAALAAVA 211
```

```
RESULT 3
US-10-156-761-10034
; Sequence 10034, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
```

;; PRIOR FILING DATE: 2001-05-30
;; PRIOR APPLICATION NUMBER: JP 2001-272697
;; PRIOR FILING DATE: 2001-08-02
;; NUMBER OF SEQ ID NOS: 15109
;; SEQ ID NO 10034
;; LENGTH: 102
;; TYPE: PRT
;; ORGANISM: Streptomyces avermitilis
US-10-156-761-10034

Query Match
Best Local Similarity 1.9%; Score 8; DB 15; Length 102;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 15 AAAVALVL 22
|||
Db 72 AAAVALVL 79

RESULT 4
US-10-080-170-577
; Sequence 577, Application US/10080170
; Publication No. US20030129601A1
; GENERAL INFORMATION:
; APPLICANT: COLE, S.T.
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
; TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
; FILE REFERENCE: 03495.0218
; CURRENT APPLICATION NUMBER: US/10/080,170
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/270,123
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 652
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 577
; LENGTH: 181
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-080-170-577

Query Match
Best Local Similarity 1.9%; Score 8; DB 16; Length 181;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 AAALAAV 18
|||
Db 16 AAALAAV 23

RESULT 5
US-10-012-542-397
; Sequence 397, Application US/10012542
; Publication No. US20030044851A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 94 Human Secreted Proteins
; FILE REFERENCE: P2029p1
; CURRENT APPLICATION NUMBER: US/10/012,542
; CURRENT FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/461,325
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-12-14
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,507
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,508
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,509
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,510
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,112
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,113

;; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22
;; NUMBER OF SEQ ID NOS: 532
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 397
;; LENGTH: 194
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-012-542-397

Query Match
Best Local Similarity 1.9%; Score 8; DB 15; Length 194;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 AALAAVA 19
|||
Db 176 AALAAVA 183

RESULT 6
US-09-813-555-1
; Sequence 1, Application US/09813555
; Patent No. US20010034455A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jun O.
; APPLICANT: Griffith, Eric C.
; TITLE OF INVENTION: Type 2 Methionine Aminopeptidase (Mecap2) Inhibitors
; TITLE OF INVENTION: and Uses Thereof
; FILE REFERENCE: 0492611-0346
; CURRENT APPLICATION NUMBER: US/09/813,555
; CURRENT FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-813-555-1

Query Match
Best Local Similarity 1.9%; Score 8; DB 9; Length 478;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 203 NAGDTTVL 210
|||
Db 235 NAGDTTVL 242

RESULT 7
US-09-813-555-2
; Sequence 2, Application US/09813555
; Patent No. US20010034455A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jun O.
; APPLICANT: Griffith, Eric C.
; APPLICANT: Su, Zhuang
; TITLE OF INVENTION: Type 2 Methionine Aminopeptidase (Mecap2) Inhibitors
; TITLE OF INVENTION: and Uses Thereof
; FILE REFERENCE: 0492611-0346
; CURRENT APPLICATION NUMBER: US/09/813,555
; CURRENT FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-813-555-2

Query Match
Best Local Similarity 1.9%; Score 8; DB 9; Length 478;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 203 NAGDTTVL 210
|||
Db 235 NAGDTTVL 242

RESULT 8

US-09-813-555-3
; Sequence 3, Application US/09813555
; Patent No. US20010034455A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jun O.
; APPLICANT: Griffith, Eric C.
; APPLICANT: Su, Zhuang
; TITLE OF INVENTION: Type 2 Methionine Aminopeptidase (MetAP2) Inhibitors
; FILE REFERENCE: 0492611-0346
; CURRENT APPLICATION NUMBER: US/09/813,555
; CURRENT FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-813-555-3

Query Match 1.9%; Score 8; DB 9; Length 478;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 203 NAGDTTVL 210
|||
Db 235 NAGDTTVL 242

RESULT 9
US-09-943-123-6
; Sequence 6, Application US/09943123
; Publication No. US20020182701A1
; GENERAL INFORMATION:
; APPLICANT: CHANG, Y-H
; APPLICANT: VETRO, J.A.
; APPLICANT: MICKA, W.S.
; TITLE OF INVENTION: Dominant Negative Variants of Methionine Aminopeptidase
; FILE REFERENCE: 16153-8007
; CURRENT APPLICATION NUMBER: US/09/943,123
; CURRENT FILING DATE: 2001-08-30
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Human dmymetap2
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (219)
; OTHER INFORMATION: May be any naturally occurring amino acid
; NAME/KEY: SITE
; LOCATION: (231)
; OTHER INFORMATION: May be any amino acid, except His
; NAME/KEY: SITE
; LOCATION: (251)
; OTHER INFORMATION: May be any naturally occurring amino acid
; NAME/KEY: SITE
; LOCATION: (262)
; OTHER INFORMATION: May be any naturally occurring amino acid
; NAME/KEY: SITE
; LOCATION: (328)
; OTHER INFORMATION: May be any naturally occurring amino acid
; NAME/KEY: SITE
; LOCATION: (331)
; OTHER INFORMATION: May be any naturally occurring amino acid
; OTHER INFORMATION: May be any naturally occurring amino acid

NAME/KEY: SITE
; LOCATION: (338)..(339)
; OTHER INFORMATION: May be any naturally occurring amino acid
; NAME/KEY: SITE
; LOCATION: (364)
; OTHER INFORMATION: May be any naturally occurring amino acid
; NAME/KEY: SITE
; LOCATION: (444)
; OTHER INFORMATION: May be any naturally occurring amino acid
; NAME/KEY: SITE
; LOCATION: (447)
; OTHER INFORMATION: May be any naturally occurring amino acid
; NAME/KEY: SITE
; LOCATION: (459)
; OTHER INFORMATION: May be any naturally occurring amino acid
US-09-943-123-6

Query Match 1.9%; Score 8; DB 10; Length 478;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 203 NAGDTTVL 210
|||
Db 235 NAGDTTVL 242

RESULT 10
US-09-943-123-7
; Sequence 7, Application US/09943123
; Publication No. US20020182701A1
; GENERAL INFORMATION:
; APPLICANT: CHANG, Y-H
; APPLICANT: VETRO, J.A.
; APPLICANT: MICKA, W.S.
; TITLE OF INVENTION: Dominant Negative Variants of Methionine Aminopeptidase
; FILE REFERENCE: 16153-8007
; CURRENT APPLICATION NUMBER: US/09/943,123
; CURRENT FILING DATE: 2001-08-30
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Mouse Metap2
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (219)
; OTHER INFORMATION: May be any naturally occurring amino acid
; NAME/KEY: SITE
; LOCATION: (231)
; OTHER INFORMATION: May be any amino acid, except His
; NAME/KEY: SITE
; LOCATION: (251)
; OTHER INFORMATION: May be any naturally occurring amino acid
; NAME/KEY: SITE
; LOCATION: (262)
; OTHER INFORMATION: May be any naturally occurring amino acid
; NAME/KEY: SITE
; LOCATION: (328)
; OTHER INFORMATION: May be any naturally occurring amino acid
; NAME/KEY: SITE
; LOCATION: (331)
; OTHER INFORMATION: May be any naturally occurring amino acid
; NAME/KEY: SITE
; LOCATION: (338)..(339)
; OTHER INFORMATION: May be any naturally occurring amino acid
; NAME/KEY: SITE
; LOCATION: (364)
; OTHER INFORMATION: May be any naturally occurring amino acid
; NAME/KEY: SITE
; LOCATION: (444)
; OTHER INFORMATION: May be any naturally occurring amino acid
; OTHER INFORMATION: May be any naturally occurring amino acid

NAME/KEY: SITE
LOCATION: (447)
OTHER INFORMATION: May be any naturally occurring amino acid
NAME/KEY: SITE
LOCATION: (459)
OTHER INFORMATION: May be any naturally occurring amino acid
US-09-943-123-7

Query Match 1.9%; Score 8; DB 10; Length 478;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 203 NAGDTTVL 210
|||||
DB 235 NAGDTTVL 242

RESULT 11
US-09-943-123-12
Sequence 12, Application US/09943123
Publication No. US20020182701A1
GENERAL INFORMATION:

APPLICANT: CHANG, Y-H
APPLICANT: VETRO, J.A.
APPLICANT: MICKA, W.S.
TITLE OF INVENTION: Dominant Negative Variants of Methionine Aminopeptidase
FILE REFERENCE: 16153-8007
CURRENT APPLICATION NUMBER: US/09/943,123
CURRENT FILING DATE: 2001-08-30
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 12
LENGTH: 478
TYPE: PRT
ORGANISM: Human MetAP2
US-09-943-123-12

Query Match 1.9%; Score 8; DB 10; Length 478;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 203 NAGDTTVL 210
|||||
DB 235 NAGDTTVL 242

RESULT 12
US-09-943-123-13
Sequence 13, Application US/09943123
Publication No. US20020182701A1
GENERAL INFORMATION:

APPLICANT: CHANG, Y-H
APPLICANT: VETRO, J.A.
APPLICANT: MICKA, W.S.
TITLE OF INVENTION: Dominant Negative Variants of Methionine Aminopeptidase
FILE REFERENCE: 16153-8007
CURRENT APPLICATION NUMBER: US/09/943,123
CURRENT FILING DATE: 2001-08-30
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 13
LENGTH: 478
TYPE: PRT
ORGANISM: Mouse MetAP2
US-09-943-123-13

Query Match 1.9%; Score 8; DB 10; Length 478;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 203 NAGDTTVL 210

DB 235 NAGDTTVL 242
|||||

RESULT 13
US-09-943-123-16
Sequence 16, Application US/09943123
Publication No. US20020182701A1
GENERAL INFORMATION:
APPLICANT: CHANG, Y-H
APPLICANT: VETRO, J.A.
APPLICANT: MICKA, W.S.
TITLE OF INVENTION: Dominant Negative Variants of Methionine Aminopeptidase
FILE REFERENCE: 16153-8007
CURRENT APPLICATION NUMBER: US/09/943,123
CURRENT FILING DATE: 2001-08-30
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 16
LENGTH: 480
TYPE: PRT
ORGANISM: Rat dnMetAP2
FEATURE:
NAME/KEY: SITE
LOCATION: (219)
OTHER INFORMATION: May be any naturally occurring amino acid
NAME/KEY: SITE
LOCATION: (231)
OTHER INFORMATION: May be any amino acid, except His
NAME/KEY: SITE
LOCATION: (251)
OTHER INFORMATION: May be any naturally occurring amino acid
NAME/KEY: SITE
LOCATION: (262)
OTHER INFORMATION: May be any naturally occurring amino acid
NAME/KEY: SITE
LOCATION: (328)
OTHER INFORMATION: May be any naturally occurring amino acid
NAME/KEY: SITE
LOCATION: (331)
OTHER INFORMATION: May be any naturally occurring amino acid
NAME/KEY: SITE
LOCATION: (338) (339)
OTHER INFORMATION: May be any naturally occurring amino acid
NAME/KEY: SITE
LOCATION: (364)
OTHER INFORMATION: May be any naturally occurring amino acid
NAME/KEY: SITE
LOCATION: (444)
OTHER INFORMATION: May be any naturally occurring amino acid
NAME/KEY: SITE
LOCATION: (447)
OTHER INFORMATION: May be any naturally occurring amino acid
NAME/KEY: SITE
LOCATION: (459)
OTHER INFORMATION: May be any naturally occurring amino acid
US-09-943-123-16

Query Match 1.9%; Score 8; DB 10; Length 480;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 203 NAGDTTVL 210
|||||
DB 235 NAGDTTVL 242

RESULT 14
US-09-943-123-17
Sequence 17, Application US/09943123
Publication No. US20020182701A1
GENERAL INFORMATION:

```

; APPLICANT: CHANG, Y-H
; APPLICANT: VETRO, J.A.
; APPLICANT: MICKA, W.S.
; TITLE OF INVENTION: Dominant Negative Variants of Methionine Aminopeptidase
; FILE REFERENCE: 16153-8007
; CURRENT APPLICATION NUMBER: US/09/943,123
; CURRENT FILING DATE: 2001-08-30
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 480
; TYPE: PRT
; ORGANISM: Rat MetAP2
US-09-943-123-17

Query Match
Best Local Similarity 1.9%; Score 8; DB 10; Length 480;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 203 NAGDTTVL 210
Db 235 NAGDTTVL 242

RESULT 15
US-10-106-698-4765
; Sequence 4765, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA0050P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 4765
; LENGTH: 500
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-4765

Query Match
Best Local Similarity 1.9%; Score 8; DB 15; Length 500;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 203 NAGDTTVL 210
Db 257 NAGDTTVL 264

RESULT 16
US-09-815-242-10147
; Sequence 10147, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: Prokaryotes
```

```

; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10147
; LENGTH: 891
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10147

Query Match
Best Local Similarity 1.9%; Score 8; DB 9; Length 891;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 RAALAAA 17
Db 36 RAALAAA 43

RESULT 17
US-09-815-242-13843
; Sequence 13843, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13843
; LENGTH: 892
; TYPE: PRT
; ORGANISM: Salmonella typhi
; FEATURE:
```



```
/ PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,949
/ PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
/ PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,953
/ PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
/ PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,950
/ PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
/ PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,947
/ PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
/ PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,964
/ PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
/ PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/056,360
/ PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
/ PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,684
/ PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
/ PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,984
/ PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
/ PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,954
/ PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
/ PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,785
/ PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
/ PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,664
/ PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
/ PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,660
/ PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
/ PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,661
/ PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
/ NUMBER OF SEQ ID NOS: 672
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 347
/ LENGTH: 24
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-883-802-347
```

```
Query Match      1.7%; Score 7; DB 11; Length 24;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      111 NLESARA 117
        |||||
Db      8 NLESARA 14
```

```
RESULT 20
US-09-864-761-44448
/ Sequence 44448, Application US/09864761
/ Patent No. US20020046763A1
/ GENERAL INFORMATION:
/ APPLICANT: Penn, Sharon G.
/ APPLICANT: Rank, David R.
/ APPLICANT: Hanzel, David K.
/ APPLICANT: Chen, Wensheng
/ TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
/ FILE REFERENCE: Aecm1ca-X-1
/ CURRENT FILING DATE: 2001-05-23
/ PRIOR APPLICATION NUMBER: US 60/180,312
/ PRIOR FILING DATE: 2000-02-04
/ PRIOR APPLICATION NUMBER: US 60/207,456
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: US 09/632,366
/ PRIOR FILING DATE: 2000-08-03
/ PRIOR APPLICATION NUMBER: GB 24263.6
/ PRIOR FILING DATE: 2000-10-04
/ PRIOR APPLICATION NUMBER: US 60/236,359
/ PRIOR FILING DATE: 2000-09-27
/ PRIOR APPLICATION NUMBER: PCT/US01/00666
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00667
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00664
/ PRIOR FILING DATE: 2001-01-30
```

```
/ PRIOR APPLICATION NUMBER: PCT/US01/00669
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00665
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00668
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00663
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00662
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00661
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00670
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: US 60/234,687
/ PRIOR FILING DATE: 2000-09-21
/ PRIOR APPLICATION NUMBER: US 09/608,408
/ PRIOR FILING DATE: 2000-06-30
/ PRIOR APPLICATION NUMBER: US 09/774,203
/ PRIOR FILING DATE: 2001-01-29
/ NUMBER OF SEQ ID NOS: 49117
/ SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
/ SEQ ID NO 44448
/ LENGTH: 37
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: MAP TO AC005039.1
/ OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.78
/ OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.79
/ OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.88
/ OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.97
/ OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1
/ OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.86
/ OTHER INFORMATION: EST_HUMAN HIT: AW512200.1, EVALU8 1.30e-01
US-09-864-761-44448
```

```
Query Match      1.7%; Score 7; DB 9; Length 37;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      66 RLESIRT 72
        |||||
Db      2 RLESIRT 8
```

```
RESULT 21
US-09-764-877-1958
/ Sequence 1958, Application US/09764877
/ Patent No. US20020147140A1
/ GENERAL INFORMATION:
/ APPLICANT: Rosen et al.
/ TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
/ FILE REFERENCE: PC005
/ CURRENT FILING DATE: 2001-01-17
/ PRIOR APPLICATION DATA REMOVED - refer to PALM or file wrapper
/ NUMBER OF SEQ ID NOS: 4031
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 1958
/ LENGTH: 65
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: SITE
/ LOCATION: (117)
/ OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
/ NAME/KEY: SITE
/ LOCATION: (63)
/ OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-877-1958
```

```
Query Match      1.7%; Score 7; DB 10; Length 65;
```

Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 232 NIMMGL 298
DB 20 NIMMGL 26

RESULT 22

US-10-156-761-13107
; Sequence 13107, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:

APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 13107
LENGTH: 117
TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-13107

Query Match 1.7%; Score 7; DB 15; Length 117;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 AAALAA 17
DB 55 AAALAA 61

RESULT 23

US-10-007-267-2
; Sequence 2, Application US/10007267
; Publication No. US20020127682A1
; GENERAL INFORMATION:

APPLICANT: Getschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
OLIGOSACCHARIDES, AND GENES ENCODING THEM
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/007,267
FILING DATE: 03-Dec-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/333,412
FILING DATE: 15-Jun-1999
APPLICATION NUMBER: 08/312,387

FILING DATE: July 7, 1994
ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-095
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:

LENGTH: 126 amino acids
TYPE: amino acid
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-007-267-2

Query Match 1.7%; Score 7; DB 14; Length 126;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 AAALAA 17
DB 13 AAALAA 19

RESULT 24

US-09-956-425-4
; Sequence 4, Application US/09956425
; Patent No. US20020045192A1
; GENERAL INFORMATION:

APPLICANT: Kriwacki, Richard
APPLICANT: Bochner, Brian
APPLICANT: Lewis, William
TITLE OF INVENTION: Aft and Hdm2 Interaction Domains and Method of Use Thereof
FILE REFERENCE: 1340/1/035
CURRENT APPLICATION NUMBER: US/09/956,425
CURRENT FILING DATE: 2001-09-19
NUMBER OF SEQ ID NOS: 25
SOFTWARE: Patentin version 3.1
SEQ ID NO 4
LENGTH: 132
TYPE: PRT
; ORGANISM: Homo sapiens
US-09-956-425-4

Query Match 1.7%; Score 7; DB 9; Length 132;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 16 AAVALV 22
DB 41 AAVALV 47

RESULT 25

US-09-867-550-150
; Sequence 150, Application US/09867550
; Patent No. US20020082206A1
; GENERAL INFORMATION:

APPLICANT: Leach, Martin D.
APPLICANT: Mehrtan, Fuad,
APPLICANT: Conley, Pamela
APPLICANT: Law, Debbie
APPLICANT: Topper, James
TITLE OF INVENTION: No. US20020082206A1 Polynucleotides from Atherogenic Cells and

FILE REFERENCE: 21402-013 (Cura-313)

CURRENT APPLICATION NUMBER: US/09/867,550
CURRENT FILING DATE: 2001-09-20
PRIOR APPLICATION NUMBER: USSN 60/208,427
PRIOR FILING DATE: 2000-05-30

NUMBER OF SEQ ID NOS: 2125
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 150
LENGTH: 132
TYPE: PRT
ORGANISM: Homo sapiens
US-09-867-550-150

Query Match 1.7%; Score 7; DB 9; Length 132;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 AALAAV 18
|||
53 AALAAV 59

RESULT 26
US-09-925-300-1688
Sequence 1688, Application US/09925300
Patent No. US20020151681A1
GENERAL INFORMATION:
APPLICANT: Craig Rosen,
APPLICANT: Steve Ruben
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA101
CURRENT APPLICATION NUMBER: US/09/925,300
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05988
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1890
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1688
LENGTH: 153
TYPE: PRT
ORGANISM: Homo sapiens
US-09-925-300-1688

Query Match 1.7%; Score 7; DB 10; Length 153;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 RAALAA 16
|||
21 RAALAA 27

RESULT 27
US-10-156-761-10928
Sequence 10928, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 10928
LENGTH: 175
TYPE: PRT

ORGANISM: Streptomyces avermitilis
US-10-156-761-10928

Query Match 1.7%; Score 7; DB 15; Length 175;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 392 AAPQSG 398
|||
79 AAPQSG 85

RESULT 28
US-10-156-761-10784
Sequence 10784, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 10784
LENGTH: 184
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-10784

Query Match 1.7%; Score 7; DB 15; Length 184;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 AAALAA 17
|||
29 AAALAA 35

RESULT 29
US-09-764-853-558
Sequence 558, Application US/09764853
Patent No. US20020090672A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: P1206
CURRENT APPLICATION NUMBER: US/09/764,853
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PAM or file wrapper
NUMBER OF SEQ ID NOS: 939
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 558
LENGTH: 203
TYPE: PRT
ORGANISM: Homo sapiens
US-09-764-853-558

Query Match 1.7%; Score 7; DB 9; Length 203;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 EAPAPV 47
|||

Db 138 EAPAPV 144

RESULT 30

US-09-738-626-6652
 ; Sequence 6652, Application US/09738626
 ; Publication No. US20020197605A1
 ; GENERAL INFORMATION:
 ; APPLICANT: NAKAGAWA, SATOSHI
 ; APPLICANT: MIZOGUCHI, HIROSHI
 ; APPLICANT: ANDO, SEIKO
 ; APPLICANT: HAYASHI, MIKIRO
 ; APPLICANT: OCHIAI, KEIKO
 ; APPLICANT: YOKOI, HARUHIKO
 ; APPLICANT: TATEISHI, NAOKO
 ; APPLICANT: SENOH, AKIHIRO
 ; APPLICANT: IKEDA, MASATO
 ; APPLICANT: OZAKI, AKIO
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-125
 ; CURRENT APPLICATION NUMBER: US/09/738, 626
 ; CURRENT FILING DATE: 2000-12-18
 ; PRIOR APPLICATION NUMBER: JP 99/377484
 ; PRIOR FILING DATE: 1999-12-16
 ; PRIOR APPLICATION NUMBER: JP 00/159162
 ; PRIOR FILING DATE: 2000-04-07
 ; PRIOR APPLICATION NUMBER: JP 00/280988
 ; PRIOR FILING DATE: 2000-08-03
 ; NUMBER OF SEQ ID NOS: 7059
 ; SOFTWARE: PatentIn ver. 3.0
 ; SEQ ID NO 6652
 ; LENGTH: 236
 ; TYPE: PR1
 ; ORGANISM: Corynebacterium glutamicum
 US-09-738-626-6652

Query Match 1.7%; Score 7; DB 10; Length 236;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 366 VEGISIA 372
 |||||
 Db 65 VEGISIA 71

RESULT 31

US-10-156-761-13873
 ; Sequence 13873, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156, 761
 ; CURRENT FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272667
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 13873
 ; LENGTH: 248
 ; TYPE: PR1
 ; ORGANISM: Streptomyces avermitilis
 US-10-156-761-13873

Query Match 1.7%; Score 7; DB 15; Length 248;

Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 233 LRQIAE 239
 |||||
 Db 119 LRQIAE 125

RESULT 32

US-10-317-806-4
 ; Sequence 4, Application US/10317806
 ; Publication No. US20030113785A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Zayed, Adel
 ; APPLICANT: Ascenzi, Robert
 ; APPLICANT: Boyes, Douglas
 ; APPLICANT: Mulpur, Rao
 ; APPLICANT: Hoffman, Neil
 ; APPLICANT: Kjettrup, Susanne
 ; APPLICANT: Davis, Keith
 ; APPLICANT: Phillips, Kenneth
 ; APPLICANT: Moesner, Jeffrey
 ; APPLICANT: Hamilton, Carol
 ; APPLICANT: Goriach, Jörn
 ; TITLE OF INVENTION: METHODS FOR THE IDENTIFICATION OF HERBICIDES AND THE MODULATION O
 ; FILE REFERENCE: Docket # 2126US
 ; CURRENT APPLICATION NUMBER: US/10/317, 806
 ; CURRENT FILING DATE: 2002-12-12
 ; NUMBER OF SEQ ID NOS: 6
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 4
 ; LENGTH: 253
 ; TYPE: PR1
 ; ORGANISM: Arabidopsis thaliana
 US-10-317-806-4

Query Match 1.7%; Score 7; DB 15; Length 253;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 403 SEAKTAS 409
 |||||
 Db 180 SEAKTAS 186

RESULT 33

US-10-156-761-14170
 ; Sequence 14170, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156, 761
 ; CURRENT FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 14170
 ; LENGTH: 262
 ; TYPE: PR1
 ; ORGANISM: Streptomyces avermitilis
 US-10-156-761-14170

Query Match 1.7%; Score 7; DB 15; Length 262;

Query Match 1.7%; Score 7; DB 15; Length 262;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 RAALAA 16
 |||||
 216 RAALAA 222

RESULT 34
 US-10-342-224-72

; Sequence 72, Application US/10342224
 ; Publication No. US20030162294A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Nathalie Verbruggen
 ; TITLE OF INVENTION: Genes Involved in Tolerance to Environmental Stress
 ; FILE REFERENCE: CN-01205
 ; CURRENT APPLICATION NUMBER: US/10/342,224
 ; CURRENT FILING DATE: 2003-01-13
 ; PRIOR APPLICATION NUMBER: US/09/762,154
 ; PRIOR FILING DATE: 2002-02-02
 ; PRIOR APPLICATION NUMBER: EP 98202634.6
 ; PRIOR FILING DATE: 1998-08-04
 ; NUMBER OF SEQ ID NOS: 123
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 72
 ; LENGTH: 271
 ; TYPE: PRT
 ; ORGANISM: Arabidopsis thaliana
 US-10-342-224-72

Query Match 1.7%; Score 7; DB 12; Length 271;
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 245 ADGVIAV 251
 |||||
 Db 136 ADGVIAV 142

RESULT 35
 US-09-815-242-10842

; Sequence 10842, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Karl L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: ELITRA 011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 10842
 ; LENGTH: 275
 ; TYPE: PRT
 ; ORGANISM: Enterococcus faecalis
 US-09-815-242-10842

Query Match 1.7%; Score 7; DB 9; Length 275;
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 319 AVTRGAK 325
 |||||
 Db 70 AVTRGAK 76

RESULT 36
 US-10-080-170-328

; Sequence 328, Application US/10080170
 ; Publication No. US20030129601A1
 ; GENERAL INFORMATION:
 ; APPLICANT: COLE, S.T.
 ; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
 ; TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
 ; FILE REFERENCE: 03495,0218
 ; CURRENT APPLICATION NUMBER: US/10/080,170
 ; CURRENT FILING DATE: 2002-06-10
 ; PRIOR APPLICATION NUMBER: 60/270,123
 ; PRIOR FILING DATE: 2001-02-22
 ; NUMBER OF SEQ ID NOS: 652
 ; SOFTWARE: Patentin Ver. 2.1
 ; SEQ ID NO 328
 ; LENGTH: 275
 ; TYPE: PRT
 ; ORGANISM: Mycobacterium leprae
 US-10-080-170-328

Query Match 1.7%; Score 7; DB 16; Length 275;
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 123 QATLAKA 129
 |||||
 Db 58 QATLAKA 64

RESULT 37
 US-10-156-761-14182

; Sequence 14182, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; CURRENT FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 14182
 ; LENGTH: 288
 ; TYPE: PRT
 ; ORGANISM: Streptomyces avermitilis
 US-10-156-761-14182

Query Match 1.7%; Score 7; DB 15; Length 288;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 AALAAA 17
DB 65 AALAAA 71

RESULT 38

US-10-156-761-8083
; Sequence 8083, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8083
; LENGTH: 300
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8083

Query Match 1.7%; Score 7; DB 15; Length 300;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 372 AGITGAK 378
DB 259 AGITGAK 265

RESULT 39

US-10-156-761-13108
; Sequence 13108, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 13108
; LENGTH: 301
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-13108

Query Match 1.7%; Score 7; DB 15; Length 301;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 244 AADGVIA 250
DB 121 AADGVIA 127

RESULT 40

US-09-533-029-28
; Sequence 28, Application US/09533029
; Publication No. US20030046723A1
; GENERAL INFORMATION:
; APPLICANT: Heard, Jacqueline
; APPLICANT: Brown, Pierre
; APPLICANT: Riechmann, Jose-Luis
; APPLICANT: Keddie, James
; APPLICANT: Pineda, Omaira
; APPLICANT: Adam, Luc
; APPLICANT: Samaha, Raymond
; APPLICANT: Zhang, James
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Pilgrim, Marsha
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Reuber, Lynne
; TITLE OF INVENTION: DISEASE-INDUCED POLYNUCLEOTIDES
; FILE REFERENCE: MBI-010
; CURRENT APPLICATION NUMBER: US/09/533,029
; CURRENT FILING DATE: 2000-03-22
; EARLIER APPLICATION NUMBER: 60/125,814
; EARLIER FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 121
; SOFTWARE: Patentm Ver. 2.1
; SEQ ID NO 28
; LENGTH: 302
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G291
US-09-533-029-28

Query Match 1.7%; Score 7; DB 11; Length 302;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 AALAAV 18
DB 284 AALAAV 290

Search completed: September 8, 2003, 14:15:02
Job time: 28 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2003, 14:06:52 ; Search time 41 Seconds
(without alignments)
2593.114 Million cell updates/sec

Title: US-09-889-756a-2

Perfect score: 412
Sequence: 1 MAFVAFKAMRAAALAAVAL.....AAPSGVOTASEAKTASEAE 412

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database :

SPREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	171	41.5	412	16	Q91T50
2	133	32.3	412	16	Q91Y66
3	100	24.3	271	2	Q51007
4	11	2.7	373	16	Q8X7E1
5	11	2.7	374	16	Q8GC84
6	11	2.7	384	16	Q8CM42
7	11	2.7	388	16	Q9AA03
8	10	2.4	124	4	Q869P9
9	10	2.4	216	4	Q81VM3
10	10	2.4	218	4	Q9BSM6
11	10	2.4	222	4	Q8TE51
12	10	2.4	222	4	Q8TAE8
13	9	2.2	444	16	Q8ACW1
14	8	1.9	107	16	Q9UQ07
15	8	1.9	125	10	Q94D46
16	8	1.9	129	16	Q32184

17	8	1.9	156	2	Q9RP11	Q9RP11 salmonella
18	8	1.9	162	10	Q8S0K1	Q8S0K1 oryza sativ
19	8	1.9	171	16	Q8XCR9	Q8XCR9 escherichia
20	8	1.9	183	16	Q8XC98	Q8XC98 escherichia
21	8	1.9	183	17	Q9YAH2	Q9YAH2 aeropyrum p
22	8	1.9	206	16	Q8P176	Q8P176 xanthomonas
23	8	1.9	206	16	Q9KZV1	Q9KZV1 streptomyce
24	8	1.9	214	16	P91115	P91115 mycobacteri
25	8	1.9	222	6	Q8SEP7	Q8SEP7 cercopitheci
26	8	1.9	238	2	Q9Z5W3	Q9Z5W3 pseudomonas
27	8	1.9	253	4	Q9EB43	Q9EB43 homo sapien
28	8	1.9	290	16	Q9S295	Q9S295 streptomyce
29	8	1.9	296	16	Q926G6	Q926G6 rhizobium m
30	8	1.9	306	16	Q9CH13	Q9CH13 lactococcus
31	8	1.9	326	10	Q9SNL7	Q9SNL7 oryza sativ
32	8	1.9	348	16	Q67350	Q67350 aquifex aeo
33	8	1.9	390	10	Q8GRS3	Q8GRS3 oryza sativ
34	8	1.9	394	2	Q9F241	Q9F241 xanthomonas
35	8	1.9	399	16	Q8P7C8	Q8P7C8 xanthomonas
36	8	1.9	408	16	Q8ZL01	Q8ZL01 salmonella
37	8	1.9	408	16	Q8Z2M0	Q8Z2M0 salmonella
38	8	1.9	416	16	Q8XBZ9	Q8XBZ9 escherichia
39	8	1.9	416	16	Q8FBW4	Q8FBW4 escherichia
40	8	1.9	442	16	Q8G7V0	Q8G7V0 bifidobacte
41	8	1.9	444	5	Q9U2V9	Q9U2V9 caenorhabdi
42	8	1.9	451	5	Q93704	Q93704 caenorhabdi
43	8	1.9	455	4	Q8NB11	Q8NB11 homo sapien
44	8	1.9	559	16	Q8FTU2	Q8FTU2 corynebacte
45	8	1.9	654	16	Q8P9J8	Q8P9J8 xanthomonas
46	8	1.9	716	16	Q8ZUJ5	Q8ZUJ5 salmonella
47	8	1.9	716	16	Q8Z0W3	Q8Z0W3 salmonella
48	8	1.9	748	16	Q9Z577	Q9Z577 streptomyce
49	8	1.9	761	16	Q9KX15	Q9KX15 streptomyce
50	8	1.9	825	5	Q8T3U6	Q8T3U6 drosophila
51	8	1.9	864	16	Q9RUB9	Q9RUB9 streptomyce
52	8	1.9	866	16	Q8EF55	Q8EF55 shewanella
53	8	1.9	871	16	Q8ZG92	Q8ZG92 yersinia pe
54	8	1.9	878	2	P74880	P74880 salmonella
55	8	1.9	891	16	Q8ZED0	Q8ZED0 salmonella
56	8	1.9	891	16	Q8FHX2	Q8FHX2 escherichia
57	8	1.9	891	16	Q8D0K2	Q8D0K2 yersinia pe
58	8	1.9	892	16	Q8ZPA5	Q8ZPA5 salmonella
59	8	1.9	892	16	Q8Z7F1	Q8Z7F1 salmonella
60	8	1.9	968	5	Q9VCB2	Q9VCB2 drosophila
61	8	1.9	979	10	Q23096	Q23096 arabidopsis
62	8	1.9	984	2	Q9XCD4	Q9XCD4 thermomono
63	8	1.9	1121	5	P90630	P90630 acanthamoeb
64	8	1.9	1291	5	Q77261	Q77261 drosophila
65	8	1.9	1300	5	Q9W5E0	Q9W5E0 drosophila
66	8	1.9	1487	5	Q8MPV5	Q8MPV5 caenorhabdi
67	8	1.9	1558	5	Q81770	Q81770 caenorhabdi
68	8	1.9	2167	5	Q76840	Q76840 caenorhabdi
69	8	1.9	2296	4	Q75042	Q75042 homo sapien
70	8	1.7	37	2	Q8LID9	Q8LID9 vibrio angu
71	7	1.7	55	6	Q8S096	Q8S096 macaca faec
72	7	1.7	55	6	Q8S090	Q8S090 pan paniscu
73	7	1.7	55	6	Q8SCA2	Q8SCA2 cercopitheci
74	7	1.7	55	6	Q8SQ99	Q8SQ99 gorilla gor
75	7	1.7	55	6	Q8SQ97	Q8SQ97 macaca arc
76	7	1.7	63	10	Q8SOA3	Q8SOA3 cercopitheci
77	7	1.7	63	10	Q9L159	Q9L159 oryza sativ
78	7	1.7	72	2	Q8VB76	Q8VB76 white spot
79	7	1.7	72	2	Q8RS17	Q8RS17 uncultured
80	7	1.7	78	2	P95455	P95455 pseudomonas
81	7	1.7	80	10	Q24280	Q24280 prunus pers
82	7	1.7	80	10	Q43750	Q43750 capsicum an
83	7	1.7	81	2	Q9APB7	Q9APB7 uncultured
84	7	1.7	87	4	Q9VB83	Q9VB83 homo sapien
85	7	1.7	91	2	Q007301	Q007301 pseudomonas
86	7	1.7	91	2	Q93M27	Q93M27 escherichia
87	7	1.7	91	15	Q8Q3B9	Q8Q3B9 human immun
88	7	1.7	91	16	Q93511	Q93511 salmonella
89	7	1.7	93	4	Q9B236	Q9B236 homo sapien

90	7	1.7	94	16	Q8P111	Q8P111 xanthomonas	163	7	1.7	203	16	Q9RK21	Q9RK21 streptomyc
91	7	1.7	96	4	Q9B337	Q9B337 homo sapien	164	7	1.7	204	2	Q8G9Q9	Q8G9Q9 helicobacte
92	7	1.7	97	4	Q9X4N8	Q9X4N8 pseudomonas	165	7	1.7	206	16	Q9X8Z0	Q9X8Z0 streptococ
93	7	1.7	97	5	P91216	P91216 caenorhabdi	166	7	1.7	207	4	Q8BVM2	Q8BVM2 homo sapien
94	7	1.7	97	15	Q8Q3C0	Q8Q3C0 human immun	167	7	1.7	208	6	Q29468	Q29468 canis famli
95	7	1.7	97	16	Q8K616	Q8K616 streptococ	168	7	1.7	208	10	Q94F04	Q94F04 lycopersico
96	7	1.7	100	11	Q8C808	Q8C808 mus musculu	169	7	1.7	209	16	Q9PLV8	Q9PLV8 campylobact
97	7	1.7	103	5	Q966Y0	Q966Y0 schistocerc	170	7	1.7	209	16	Q9A577	Q9A577 caulobacter
98	7	1.7	103	5	Q9SPM5	Q9SPM5 schistocerc	171	7	1.7	211	2	Q931I8	Q931I8 thibacilli
99	7	1.7	104	5	Q9SPM4	Q9SPM4 schistocerc	172	7	1.7	211	2	Q82867	Q82867 rhodovuln
100	7	1.7	104	13	Q9W668	Q9W668 xenopus lae	173	7	1.7	212	16	Q8YR05	Q8YR05 anabena sp
101	7	1.7	106	9	Q9T111	Q9T111 bacterioph	174	7	1.7	212	17	Q9UX14	Q9UX14 pyrococcus
102	7	1.7	107	5	Q966Y1	Q966Y1 schistocerc	175	7	1.7	213	10	Q9Z081	Q9Z081 arabidopsis
103	7	1.7	113	2	Q9X9S3	Q9X9S3 streptomyc	176	7	1.7	214	16	Q8UDC6	Q8UDC6 agrobacteri
104	7	1.7	124	16	Q910T5	Q910T5 streptomyc	177	7	1.7	215	2	Q934G2	Q934G2 mycobacteri
105	7	1.7	125	10	Q910T3	Q910T3 arabidopsis	178	7	1.7	217	2	Q8KM09	Q8KM09 ruegeria sp
106	7	1.7	131	16	Q8YH06	Q8YH06 bruceella me	179	7	1.7	217	10	Q39826	Q39826 glycine max
107	7	1.7	131	16	Q8G015	Q8G015 bruceella su	180	7	1.7	221	6	Q9SK08	Q9SK08 sus scrofa
108	7	1.7	132	4	Q96B52	Q96B52 homo sapien	181	7	1.7	223	2	Q87773	Q87773 lactobacilli
109	7	1.7	132	4	Q13195	Q13195 homo sapien	182	7	1.7	223	16	Q92547	Q92547 streptomyc
110	7	1.7	132	10	Q9AY50	Q9AY50 chlamydomon	183	7	1.7	225	10	Q949B2	Q949B2 oryza sativ
111	7	1.7	134	2	Q48533	Q48533 leprothrix	184	7	1.7	225	16	Q8FMR8	Q8FMR8 corynebacte
112	7	1.7	135	16	Q9A6E3	Q9A6E3 caulobacter	185	7	1.7	228	8	Q9B661	Q9B661 baiganea ru
113	7	1.7	138	2	Q8KLF1	Q8KLF1 rhizobium e	186	7	1.7	229	2	Q938D9	Q938D9 mycobacteri
114	7	1.7	141	10	Q9AX11	Q9AX11 oryza sativ	187	7	1.7	229	4	Q81Y89	Q81Y89 homo sapien
115	7	1.7	142	2	Q46987	Q46987 escherichia	188	7	1.7	230	16	Q9A513	Q9A513 caulobacter
116	7	1.7	145	16	Q8UB04	Q8UB04 agrobacteri	189	7	1.7	230	16	Q9K3L3	Q9K3L3 streptomyc
117	7	1.7	147	16	O07754	O07754 mycobacteri	190	7	1.7	231	16	P94516	P94516 bacillus su
118	7	1.7	149	10	Q94102	Q94102 oryza sativ	191	7	1.7	233	5	Q9SW10	Q9SW10 crassostrea
119	7	1.7	150	5	Q8T918	Q8T918 drosophila	192	7	1.7	233	10	Q9M224	Q9M224 arabidopsis
120	7	1.7	152	2	Q9XDE5	Q9XDE5 bruceella me	193	7	1.7	234	4	Q9BRM1	Q9BRM1 homo sapien
121	7	1.7	152	5	Q9WEC5	Q9WEC5 drosophila	194	7	1.7	235	13	Q98TY8	Q98TY8 oreochromis
122	7	1.7	152	16	Q9ABU8	Q9ABU8 caulobacter	195	7	1.7	235	13	Q9PTD4	Q9PTD4 ictalurys p
123	7	1.7	154	16	Q8ZAX2	Q8ZAX2 yersinia pe	196	7	1.7	236	4	Q8BZ35	Q8BZ35 homo sapien
124	7	1.7	155	10	Q9FT11	Q9FT11 capsicum an	197	7	1.7	240	16	Q8YBR4	Q8YBR4 bruceella me
125	7	1.7	155	10	Q949A8	Q949A8 oryza sativ	198	7	1.7	241	11	Q91ZP6	Q91ZP6 mus musculu
126	7	1.7	155	16	Q8B3G2	Q8B3G2 xanthomonas	199	7	1.7	241	11	Q8BVC5	Q8BVC5 mus musculu
127	7	1.7	156	10	Q94ED8	Q94ED8 oryza sativ	200	7	1.7	242	16	Q9PCD9	Q9PCD9 xyella fas
128	7	1.7	157	17	Q56919	Q56919 pyrococcus	201	7	1.7	243	10	Q82640	Q82640 arabidopsis
129	7	1.7	158	16	Q8UG86	Q8UG86 agrobacteri	202	7	1.7	243	10	Q945W2	Q945W2 oryza sativ
130	7	1.7	163	2	Q8GDE8	Q8GDE8 wolbachia e	203	7	1.7	243	10	Q945X1	Q945X1 oryza sativ
131	7	1.7	163	5	Q10054	Q10054 caenorhabdi	204	7	1.7	244	16	Q8ZFT5	Q8ZFT5 yersinia pe
132	7	1.7	165	2	P77910	P77910 neisseria g	205	7	1.7	245	16	Q8EANA	Q8EANA shewanella
133	7	1.7	167	16	Q8R7F5	Q8R7F5 thermoaer	206	7	1.7	249	16	Q9POM9	Q9POM9 ureaplasma
134	7	1.7	168	10	Q9S948	Q9S948 lycopersico	207	7	1.7	253	10	Q9XFM5	Q9XFM5 arabidopsis
135	7	1.7	168	16	Q8XGZ0	Q8XGZ0 yersinia pe	208	7	1.7	254	11	Q8C3C7	Q8C3C7 mus musculu
136	7	1.7	169	10	Q9S950	Q9S950 lycopersico	209	7	1.7	254	17	Q8U2D4	Q8U2D4 pyrococcus
137	7	1.7	169	10	Q9S947	Q9S947 lycopersico	210	7	1.7	258	10	Q9AUU0	Q9AUU0 oryza sativ
138	7	1.7	169	10	Q9S885	Q9S885 itegaria an	211	7	1.7	258	16	Q92MX0	Q92MX0 rhizobium m
139	7	1.7	169	16	Q8D1P3	Q8D1P3 yersinia pe	212	7	1.7	259	2	Q54168	Q54168 streptomyc
140	7	1.7	171	16	Q8DBP1	Q8DBP1 vibrio vuln	213	7	1.7	260	5	Q95T44	Q95T44 drosophila
141	7	1.7	172	10	Q41119	Q41119 phaseolus v	214	7	1.7	260	16	Q8F545	Q8F545 corynebacte
142	7	1.7	172	16	Q8DHM7	Q8DHM7 synecococ	215	7	1.7	264	5	Q95Q16	Q95Q16 caenorhabdi
143	7	1.7	173	2	Q9F4J9	Q9F4J9 streptococ	216	7	1.7	264	16	Q98K02	Q98K02 rhizobium l
144	7	1.7	173	4	Q13399	Q13399 homo sapien	217	7	1.7	265	16	Q91L61	Q91L61 streptomyc
145	7	1.7	173	4	Q8N726	Q8N726 homo sapien	218	7	1.7	265	16	Q8PDP4	Q8PDP4 xanthomonas
146	7	1.7	174	17	Q8R021	Q8R021 methanosarc	219	7	1.7	266	2	Q33966	Q33966 saccharopol
147	7	1.7	179	4	Q16360	Q16360 homo sapien	220	7	1.7	267	16	Q9RWS4	Q9RWS4 deinococcus
148	7	1.7	179	16	Q8FLZ2	Q8FLZ2 corynebacte	221	7	1.7	268	12	Q91T10	Q91T10 lupula herp
149	7	1.7	180	16	Q9RZ30	Q9RZ30 deinococcus	222	7	1.7	269	5	Q9NCF3	Q9NCF3 periplaneta
150	7	1.7	181	17	Q82615	Q82615 methanobact	223	7	1.7	269	5	Q9NCF6	Q9NCF6 mactocermes
151	7	1.7	182	2	Q93ET2	Q93ET2 cowdria rum	224	7	1.7	270	4	Q96CE9	Q96CE9 corynebacte
152	7	1.7	185	16	Q98717	Q98717 rhizobium l	225	7	1.7	270	5	Q9NCF1	Q9NCF1 cryptocercu
153	7	1.7	185	16	Q8FLX9	Q8FLX9 corynebacte	226	7	1.7	270	5	Q9NCF5	Q9NCF5 mactocermes
154	7	1.7	185	16	Q8RTU3	Q8RTU3 melanosarc	227	7	1.7	270	5	Q9NCF0	Q9NCF0 cryptocercu
155	7	1.7	187	8	Q8HCP3	Q8HCP3 oryza sativ	228	7	1.7	270	5	Q9NCF8	Q9NCF8 neotermes k
156	7	1.7	191	5	Q9N936	Q9N936 leishmania	229	7	1.7	270	5	Q9NCF5	Q9NCF5 polyphaga a
157	7	1.7	191	6	Q28583	Q28583 ovis aries	230	7	1.7	270	5	Q9NCF9	Q9NCF9 cryiocercu
158	7	1.7	193	5	Q20186	Q20186 caenorhabdi	231	7	1.7	270	5	Q9NCF6	Q9NCF6 polyphaga a
159	7	1.7	197	6	Q28584	Q28584 ovis aries	232	7	1.7	270	5	Q9NCF4	Q9NCF4 blattella g
160	7	1.7	198	16	Q9KY89	Q9KY89 streptomyc	233	7	1.7	270	5	Q9NCF7	Q9NCF7 hodotermops
161	7	1.7	202	16	Q9AA57	Q9AA57 caulobacter	234	7	1.7	271	11	Q9R1X2	Q9R1X2 mus musculu
162	7	1.7	203	4	Q8NCR1	Q8NCR1 homo sapien	235	7	1.7	271	11	Q9WV22	Q9WV22 rattus norv

236	7	1.7	272	5	Q81PA7	Q81pa7 drosophila	309	7	1.7	353	2	P70901	P70901 borrelia he
237	7	1.7	273	12	Q841V5	Q841V5 cotesia con	310	7	1.7	354	10	Q9SYZ8	Q9SYZ8 arabidopsis
238	7	1.7	273	12	Q9YID2	Q9Yid2 cotesia con	311	7	1.7	355	2	Q9ANE3	Q9ANE3 bradyrhizob
239	7	1.7	273	15	Q8UE53	Q8ue53 agrobacteri	312	7	1.7	355	10	Q42853	Q42853 hordeum vul
240	7	1.7	274	5	Q81PA6	Q81pa6 drosophila	313	7	1.7	355	17	Q97W47	Q97W47 sulfobolus
241	7	1.7	275	5	Q8ETX3	Q8etx3 encephalito	314	7	1.7	357	10	Q8H385	Q8H385 oryza sativ
242	7	1.7	275	16	Q9X793	Q9X793 mycobacteri	315	7	1.7	358	2	P70899	P70899 borrelia he
243	7	1.7	275	16	Q98LT3	Q98lt3 rhizobium 1	316	7	1.7	358	16	Q9RCF4	Q9RCF4 vibrio chol
244	7	1.7	277	17	Q8XRS0	Q8xrs0 ralsstonia s	317	7	1.7	360	2	P70905	P70905 borrelia he
245	7	1.7	277	17	Q9HP12	Q9hp12 halobacteri	318	7	1.7	361	10	Q9LZ09	Q9LZ09 arabidopsis
246	7	1.7	280	5	Q81PA5	Q81pa5 drosophila	319	7	1.7	363	12	Q9OJ57	Q9OJ57 human herpe
247	7	1.7	280	11	Q9CRV1	Q9crv1 mus musculu	320	7	1.7	364	10	Q42854	Q42854 hordeum vul
248	7	1.7	281	10	Q8LXK0	Q8lxx0 zea mays (m	321	7	1.7	366	5	Q9NB12	Q9NB12 drosophila
249	7	1.7	281	10	Q8GZ05	Q8gz05 zea mays (m	322	7	1.7	366	5	Q9NB16	Q9NB16 drosophila
250	7	1.7	283	16	Q9PCE4	Q9pce4 xyella fas	323	7	1.7	366	5	Q9NB14	Q9NB14 drosophila
251	7	1.7	284	10	Q9AU05	Q9au05 chlamydomon	324	7	1.7	367	2	Q8RTE5	Q8RTE5 campylobact
252	7	1.7	286	16	Q9HWA9	Q9hw49 pseudomonas	325	7	1.7	367	5	Q81082	Q81082 campylobact
253	7	1.7	287	10	Q949C0	Q949c0 oryza sativ	326	7	1.7	367	16	Q9PID5	Q9PID5 campylobact
254	7	1.7	288	10	Q9SNA3	Q9sn83 arabidopsis	327	7	1.7	367	16	Q69922	Q69922 streptomyce
255	7	1.7	289	10	Q8SBA6	Q8sb46 oryza sativ	328	7	1.7	369	16	Q9AJK7	Q9AJK7 caulobacter
256	7	1.7	291	17	Q8ZKJ6	Q8zj16 pyrobaculum	329	7	1.7	370	5	Q8IIZ9	Q8IIZ9 plasmodium
257	7	1.7	292	16	Q8CTSO	Q8ct50 staphylococ	330	7	1.7	370	10	Q9IGT5	Q9IGT5 oryza sativ
258	7	1.7	293	16	P73482	P73482 synchocyst	331	7	1.7	371	3	Q01787	Q01787 planerociae
259	7	1.7	294	16	Q8PP88	Q8pp58 xanthomonas	332	7	1.7	372	8	Q953V6	Q953V6 lampyropelti
260	7	1.7	295	16	Q92BG7	Q92bg7 listeria in	333	7	1.7	372	8	Q953R5	Q953R5 lampyropelti
261	7	1.7	295	16	Q8Y6Y4	Q8y6y4 listeria mo	334	7	1.7	372	8	Q953T0	Q953T0 lampyropelti
262	7	1.7	296	16	Q98K10	Q98k10 rhizobium 1	335	7	1.7	374	10	Q9SYZ7	Q9SYZ7 arabidopsis
263	7	1.7	296	16	Q8PH26	Q8ph26 xanthomonas	336	7	1.7	375	16	Q8PR16	Q8PR16 xanthomonas
264	7	1.7	298	16	Q8PMW5	Q8pmw5 xanthomonas	337	7	1.7	376	16	Q92WN8	Q92WN8 rhizobium m
265	7	1.7	301	2	Q926F0	Q926f0 mycobacteri	338	7	1.7	377	5	Q81D08	Q81D08 plasmodium
266	7	1.7	302	10	Q8L905	Q8l915 arabidopsis	339	7	1.7	378	10	Q8W2W3	Q8W2W3 bovine herp
267	7	1.7	305	5	Q81PA8	Q81pa8 drosophila	340	7	1.7	378	12	Q65822	Q65822 bovine herp
268	7	1.7	307	10	Q9FJU6	Q9fj16 arabidopsis	341	7	1.7	379	5	Q81AP3	Q81AP3 plasmodium
269	7	1.7	307	16	Q984P8	Q984p8 rhizobium 1	342	7	1.7	379	16	Q8CY39	Q8CY39 brucella su
270	7	1.7	308	16	Q9KNCS	Q9knc5 vibrio chol	343	7	1.7	380	16	Q98180	Q98180 rhizobium 1
271	7	1.7	309	16	Q8NS10	Q8ns10 corynebacte	344	7	1.7	381	5	Q813D9	Q813D9 plasmodium
272	7	1.7	309	10	Q8G158	Q8gy58 arabidopsis	345	7	1.7	381	11	Q8BNF2	Q8BNF2 mus musculu
273	7	1.7	310	2	Q9FDD6	Q9fda6 selemomonas	346	7	1.7	382	16	Q8E8H1	Q8E8H1 shewanella
274	7	1.7	310	10	Q8S9P6	Q8s9p6 oryza sativ	347	7	1.7	385	5	Q25331	Q25331 leishmania
275	7	1.7	312	4	Q95286	Q95286 homo sapien	348	7	1.7	385	16	Q00745	Q00745 mycobacteri
276	7	1.7	312	16	Q8EP97	Q8ep97 shewanella	349	7	1.7	385	16	Q8XT04	Q8XT04 ralsstonia s
277	7	1.7	315	16	Q9K0A9	Q9k0a9 neisseria m	350	7	1.7	386	10	Q8STB8	Q8STB8 oryza sativ
278	7	1.7	315	16	Q9JVB5	Q9jvb5 neisseria m	351	7	1.7	386	10	Q9F0H9	Q9F0H9 zea mays (m
279	7	1.7	315	16	Q9RUS7	Q9rls7 streptomyce	352	7	1.7	388	16	Q8XVV1	Q8XVV1 ralsstonia s
280	7	1.7	316	16	Q92TA6	Q92ta6 rhizobium m	353	7	1.7	389	2	Q9ZNC9	Q9ZNC9 pseudomonas
281	7	1.7	317	10	Q93WY7	Q93wy7 nicotiana t	354	7	1.7	390	2	Q9APV7	Q9APV7 treponema d
282	7	1.7	317	10	Q93WY8	Q93wy8 nicotiana t	355	7	1.7	392	2	Q9F3Y3	Q9F3Y3 pasteurella
283	7	1.7	319	16	Q8ZAT8	Q8zac8 yerisina pe	356	7	1.7	392	16	Q8PEB7	Q8PEB7 xanthomonas
284	7	1.7	319	16	Q38817	Q38817 arabidopsis	357	7	1.7	393	16	Q50993	Q50993 neisseria g
285	7	1.7	320	16	Q8NPV3	Q8npv3 corynebacte	358	7	1.7	393	16	Q9JZP6	Q9JZP6 streptomyce
286	7	1.7	320	17	Q8TL26	Q8tl26 methanosarc	359	7	1.7	393	16	Q50532	Q50532 streptomyce
287	7	1.7	324	5	Q9XTY9	Q9xtv9 caenorhabdi	360	7	1.7	394	5	Q9NJS6	Q9NJS6 aplysia kur
288	7	1.7	324	16	Q9PH13	Q9ph13 xyella fas	361	7	1.7	394	5	Q9NHP3	Q9NHP3 aplysia cal
289	7	1.7	325	10	Q43751	Q43751 capsitum an	362	7	1.7	395	16	Q9ZBN5	Q9ZBN5 streptomyce
290	7	1.7	325	11	Q8BRB6	Q8brb6 mus musculu	363	7	1.7	396	2	Q9ZK25	Q9ZK25 pseudomonas
291	7	1.7	330	2	Q8GRE6	Q8gre6 borrelia du	364	7	1.7	396	16	Q9RG60	Q9RG60 pseudomonas
292	7	1.7	330	16	Q9PUH4	Q9puh4 chlamydia m	365	7	1.7	398	2	Q9EXN8	Q9EXN8 escherichia
293	7	1.7	331	16	Q9ACU4	Q9acu4 streptomyce	366	7	1.7	398	2	Q9Z1B3	Q9Z1B3 proteus mir
294	7	1.7	332	4	Q8N2U1	Q8n2u1 homo sapien	367	7	1.7	398	16	Q8Y3G9	Q8Y3G9 ralsstonia s
295	7	1.7	332	11	Q8COK5	Q8cok5 mus musculu	368	7	1.7	398	16	Q8PHS0	Q8PHS0 xanthomonas
296	7	1.7	333	16	Q8FUD0	Q8fud0 corynebacte	369	7	1.7	398	17	Q30103	Q30103 archaeoglob
297	7	1.7	334	16	Q8YU05	Q8yu05 ralsstonia s	370	7	1.7	399	2	Q87935	Q87935 burkholderi
298	7	1.7	336	10	Q8S7Y4	Q8s7y4 oryza sativ	371	7	1.7	399	16	Q8RR04	Q8RR04 corynebacte
299	7	1.7	336	16	Q92VM9	Q92vm9 rhizobium m	372	7	1.7	400	2	Q9ZAI1	Q9ZAI1 pasteurella
300	7	1.7	339	16	Q8XXA2	Q8xxa2 ralsstonia s	373	7	1.7	400	16	Q9R758	Q9R758 pasteurella
301	7	1.7	340	17	Q8T1M4	Q8t1m4 methanosarc	374	7	1.7	400	16	Q8G7Y2	Q8G7Y2 bifidobacte
302	7	1.7	341	16	Q98GK2	Q98gk2 rhizobium 1	375	7	1.7	401	16	Q8CUM3	Q8CUM3 streptomyce
303	7	1.7	345	2	Q8GRD4	Q8grd4 borrelia du	376	7	1.7	402	16	Q98813	Q98813 rhizobium 1
304	7	1.7	346	16	Q9AKS24	Q9aks24 ralsstonia s	377	7	1.7	402	16	Q69865	Q69865 streptomyce
305	7	1.7	348	16	Q9AK41	Q9ak41 streptomyce	378	7	1.7	402	16	Q8CNA1	Q8CNA1 staphylococ
306	7	1.7	348	17	Q8TZR7	Q8tzr7 pyrococcus	379	7	1.7	403	16	Q9UT2	Q9UT2 neisseria m
307	7	1.7	349	2	Q9F811	Q9f811 ewhina amy	380	7	1.7	404	16	Q06263	Q06263 mycobacteri
308	7	1.7	350	16	Q9X8V8	Q9x8v8 streptomyce	381	7	1.7	404	16	Q8X264	Q8X264 ralsstonia s

382	1.7	405	2	Q934L1	Q93411 vibrio salm	455	1.7	487	4	Q95172	Q95172 homo sapien
383	1.7	405	2	Q52225	Q52225 agrobacteri	456	1.7	488	10	Q22298	Q22298 citrus sine
384	1.7	406	16	Q8DB09	Q8db09 vibrio vuln	457	1.7	488	11	Q8K3V8	Q8K3V8 rattus norv
385	1.7	407	2	Q92619	Q92619 anaplasm m	458	1.7	489	10	Q42872	Q42872 lycopersico
386	1.7	407	10	Q9S7J3	Q9s7j3 welitschia	459	1.7	489	10	Q49296	Q49296 arabidopsis
387	1.7	409	2	Q44092	Q44092 anaplasm m	460	1.7	489	10	Q93W21	Q93w21 nicotiana t
388	1.7	409	10	Q8H653	Q8h653 oryza sativ	461	1.7	490	16	Q8XW69	Q8xw69 ralsconia s
389	1.7	409	16	Q69974	Q69974 streptomyc	462	1.7	492	10	Q9CAC1	Q9cac1 arabidopsis
390	1.7	410	2	Q8RSK4	Q8rsk4 uncultured	463	1.7	492	10	Q23697	Q23697 arabidopsis
391	1.7	410	2	Q9R965	Q9r965 anaplasm m	464	1.7	493	10	Q23696	Q23696 arabidopsis
392	1.7	410	16	Q92JW7	Q92jw7 rhizobium m	465	1.7	493	10	Q9AYV9	Q9ayv9 atriplex le
393	1.7	411	2	Q9RH56	Q9rh56 bradyrhizob	466	1.7	493	10	Q9SVJ3	Q9svj3 arabidopsis
394	1.7	411	17	Q8PYR2	Q8pyr2 methanosaic	467	1.7	494	10	Q9SVJ4	Q9svj4 arabidopsis
395	1.7	417	16	Q91478	Q91478 pseudomonas	468	1.7	494	10	Q9X1Y8	Q9x1y8 populus alb
396	1.7	417	16	Q9RY67	Q9ry67 deinococcus	469	1.7	494	10	Q40763	Q40763 populus alb
397	1.7	418	17	Q9YF00	Q9yfg0 aeropyrum p	470	1.7	494	10	Q9AVT5	Q9avt5 populus alb
398	1.7	420	17	Q9HNG8	Q9hng8 halobacteri	471	1.7	496	10	Q81403	Q81403 fragaria an
399	1.7	422	16	Q8PRB8	Q8prb8 xanthomonas	472	1.7	496	10	Q92T50	Q92t50 fragaria an
400	1.7	423	2	Q92620	Q92620 anaplasm m	473	1.7	496	10	Q9ZTL0	Q9ztl0 fragaria an
401	1.7	431	17	Q9H161	Q9h161 thermoplasm	474	1.7	496	10	Q9S8T6	Q9sb6 fragaria an
402	1.7	433	5	Q96819	Q96819 trypanosoma	475	1.7	497	5	Q9GRR2	Q9grr2 leishmania
403	1.7	433	12	Q9WT62	Q9wt62 human herpe	476	1.7	497	10	Q82473	Q82473 lycopersico
404	1.7	435	5	Q817C8	Q817c8 ciona intes	477	1.7	497	10	Q96546	Q96546 capsicum an
405	1.7	436	2	Q61566	Q61566 streptomyc	478	1.7	497	10	Q9SML6	Q9sm6 capsicum an
406	1.7	436	16	Q8XAF6	Q8xaf6 escherichia	479	1.7	497	10	P94114	P94114 prunus pers
407	1.7	436	16	Q8XAF6	Q8xaf6 escherichia	480	1.7	498	4	Q9NPE8	Q9npe8 homo sapien
408	1.7	437	2	Q9F1V8	Q9f1v8 streptomyc	481	1.7	499	10	Q8LO92	Q8l92 oryza sativ
409	1.7	440	12	Q65588	Q65588 bovine herp	482	1.7	499	10	Q8KY30	Q8ky30 streptomyc
410	1.7	441	10	Q94C38	Q94c38 arabidopsis	483	1.7	501	10	Q64949	Q64949 arabidopsis
411	1.7	441	16	Q9RZV2	Q9rvz2 deinococcus	484	1.7	501	10	Q9SRK3	Q9srk3 arabidopsis
412	1.7	443	11	Q91ZD7	Q91zd7 rattus norv	485	1.7	502	16	Q8ZGP5	Q8zgp5 yerania pe
413	1.7	444	5	Q9V839	Q9v839 drosophila	486	1.7	504	10	Q9LDF0	Q9ldf0 oryza sativ
414	1.7	444	16	Q8XZ44	Q8xz44 ralsconia s	487	1.7	505	10	Q22297	Q22297 citrus sine
415	1.7	447	5	Q9NCE2	Q9nce2 panesthia c	488	1.7	506	16	Q9SSU7	Q9ssu7 distum sativ
416	1.7	447	5	Q81FUE	Q81fue mastoetermes	489	1.7	506	16	Q8PSW8	Q8psw8 xanthomonas
417	1.7	447	5	Q81FUE	Q81fue mastoetermes	490	1.7	507	10	Q9ATV1	Q9atv1 lolium rigi
418	1.7	447	16	Q9KMX3	Q9kmx3 vibrio chol	491	1.7	507	16	Q8CK16	Q8ck16 streptomyc
419	1.7	447	16	Q9K935	Q9k935 bacillus ha	492	1.7	508	10	Q43149	Q43149 sambucus ni
420	1.7	448	2	Q93HP4	Q93hp4 streptomyc	493	1.7	508	16	Q86591	Q86591 streptomyc
421	1.7	448	5	Q9BMC7	Q9bmc7 coproetermes	494	1.7	509	16	Q8Y3F0	Q8y3f0 ralsconia s
422	1.7	448	5	Q77045	Q77045 nasutiterme	495	1.7	510	5	Q44018	Q44018 leishmania
423	1.7	448	5	Q77044	Q77044 nasutiterme	496	1.7	510	10	Q9ATV2	Q9atv2 lolium rigi
424	1.7	448	5	Q81FUE	Q81fue mastoetermes	497	1.7	510	10	Q42875	Q42875 lycopersico
425	1.7	448	5	Q81FUE	Q81fue mastoetermes	498	1.7	510	10	Q64401	Q64401 pinus radia
426	1.7	449	5	Q81FUE	Q81fue mastoetermes	499	1.7	511	10	Q9LE72	Q9le72 oryza sativ
427	1.7	449	16	Q8RCG1	Q8rcg1 thermomater	500	1.7	514	4	Q9BWV3	Q9bwv3 homo sapien
428	1.7	450	5	Q9NCE3	Q9nce3 panesthia c	501	1.7	515	10	Q64402	Q64402 pinus radia
429	1.7	456	2	Q9AF07	Q9af07 frankia sp.	502	1.7	516	10	Q81416	Q81416 arabidopsis
430	1.7	456	5	Q9N9U6	Q9n9u6 leishmania	503	1.7	519	5	Q8MRX3	Q8mrx3 drosophila
431	1.7	460	10	Q8L1Y8	Q8l1y8 oryza sativ	504	1.7	525	16	Q8RCG5	Q8rcg5 fusobacteri
432	1.7	461	5	Q8T7Y2	Q8t7y2 bombyx mori	505	1.7	529	2	Q50747	Q50747 mycobacteri
433	1.7	464	10	Q9AHY3	Q9ahy3 oryza sativ	506	1.7	530	3	Q14430	Q14430 cryphonectr
434	1.7	465	4	Q96GY4	Q96gy4 homo sapien	507	1.7	530	16	Q05813	Q05813 mycobacteri
435	1.7	469	5	Q9Y0W2	Q9y0w2 cherax quad	508	1.7	531	16	Q98D82	Q98d82 rhizobium l
436	1.7	470	5	Q9V3E1	Q9v3e1 drosophila	509	1.7	532	2	Q9F7M5	Q9f7m5 uncultured
437	1.7	470	5	Q8SVF2	Q8svf2 drosophila	510	1.7	533	16	Q53903	Q53903 streptomyc
438	1.7	471	10	Q8LAG1	Q8lag1 arabidopsis	511	1.7	535	16	Q9CBT1	Q9cbt1 mycobacteri
439	1.7	471	10	Q8VZES	Q8vze5 arabidopsis	512	1.7	536	2	Q8BRC1	Q8br1 comanona a
440	1.7	473	2	Q9ROE6	Q9roe6 cellulomona	513	1.7	538	5	Q9Y115	Q9y115 drosophila
441	1.7	477	4	Q8NAT2	Q8nat2 homo sapien	514	1.7	539	16	Q9RK08	Q9rk08 streptomyc
442	1.7	479	10	Q04972	Q04972 lycopersico	515	1.7	540	17	Q8TPJ7	Q8tpj7 methanosaic
443	1.7	479	10	Q9SUS0	Q9sus0 arabidopsis	516	1.7	541	10	Q9SGR6	Q9sg6 arabidopsis
444	1.7	480	10	Q9S290	Q9s290 arabidopsis	517	1.7	541	10	Q93ZJ4	Q93zj4 arabidopsis
445	1.7	481	10	Q9ZVZ7	Q9zvz7 arabidopsis	518	1.7	541	10	Q8L790	Q8l790 arabidopsis
446	1.7	482	16	Q8CJX6	Q8cjx6 streptomyc	519	1.7	543	11	Q8BNH8	Q8bnh8 mus musculu
447	1.7	484	2	Q9LAV5	Q9lav5 thermomater	520	1.7	543	16	Q69831	Q69831 streptomyc
448	1.7	484	10	Q9C9H5	Q9c9h5 arabidopsis	521	1.7	546	16	Q986W2	Q986w2 rhizobium l
449	1.7	484	10	Q8LDB8	Q8ldb8 arabidopsis	522	1.7	551	16	Q9HUD4	Q9hud4 pseudomonas
450	1.7	485	10	Q96547	Q96547 capsicum an	523	1.7	552	3	Q96TJ9	Q96tj9 utromyces fa
451	1.7	486	10	Q41012	Q41012 pisum sativ	524	1.7	555	11	Q8C3W8	Q8c3w8 mus musculu
452	1.7	486	10	Q9FKQ1	Q9fkq1 arabidopsis	525	1.7	555	11	Q8C3C6	Q8c3c6 mus musculu
453	1.7	487	4	Q9UGA3	Q9uga3 homo sapien	526	1.7	555	11	Q8BH92	Q8bh92 mus musculu
454	1.7	487	4	Q9UNW1	Q9unw1 homo sapien	527	1.7	559	11	Q8R136	Q8r136 mus musculu

528	7	1.7	559	17	Q9HRT5	Q9HRT5 halobacteri	601	7	1.7	773	16	Q9L1R8	Q9L1R8 streptomyc
529	7	1.7	571	10	Q04250	Q04250 arabidopsis	602	7	1.7	777	5	O61366	O61366 drosophila
530	7	1.7	572	3	Q42728	Q42728 rhodospirid	603	7	1.7	782	16	Q988Y2	Q988Y2 rhizobium 1
531	7	1.7	574	16	Q91246	Q91246 pseudomonas	604	7	1.7	786	5	O8T107	O8T107 bombax mori
532	7	1.7	581	2	O51863	O51863 escherichia	605	7	1.7	786	5	Q24014	Q24014 dictyosteli
533	7	1.7	581	2	Q9S6H9	Q9S6H9 escherichia	606	7	1.7	789	16	Q8UGX6	Q8UGX6 agrobacteri
534	7	1.7	582	5	O62146	O62146 caenorhabdi	607	7	1.7	802	4	Q9BQH6	Q9BQH6 homo sapien
535	7	1.7	582	16	O67141	O67141 aquifex aeo	608	7	1.7	813	2	O8GFH6	Q8GFH3 streptomyc
536	7	1.7	584	17	Q9HON3	Q9HON3 halobacteri	609	7	1.7	827	10	O8S5J7	O8S5J7 oryza sativ
537	7	1.7	586	4	Q8NA76	Q8NA76 homo sapien	610	7	1.7	840	2	O44003	Q44003 ralstonia m
538	7	1.7	586	4	O81WD5	O81WD5 homo sapien	611	7	1.7	850	16	Q8XZD5	Q8XZD5 ralstonia s
539	7	1.7	586	10	Q949C3	Q949C3 oryza sativ	612	7	1.7	859	15	Q8O850	Q8O850 human immun
540	7	1.7	587	16	Q8DCU4	Q8DCU4 vibrio vuln	613	7	1.7	860	2	Q9L428	Q9L428 bordetella
541	7	1.7	592	16	Q8PGF6	Q8PGF6 xanthomonas	614	7	1.7	870	5	O8MSY6	O8MSY6 drosophila
542	7	1.7	594	3	Q9P6E2	Q9P6E2 neurospora	615	7	1.7	875	16	O8RB43	O8RB43 thermoaer
543	7	1.7	594	16	Q8PEF6	Q8PEF6 xanthomonas	616	7	1.7	884	10	Q94FE0	Q94FE0 dichantheli
544	7	1.7	595	4	Q9ULP9	Q9ULP9 homo sapien	617	7	1.7	884	10	Q8GV57	Q8GV57 oryza sativ
545	7	1.7	597	5	O02137	O02137 caenorhabdi	618	7	1.7	887	1	O50743	Q50743 methanobact
546	7	1.7	600	16	Q9H2M1	Q9H2M1 pseudomonas	619	7	1.7	891	11	O8VHG2	Q8VHG2 mus musculu
547	7	1.7	600	16	Q8G5G9	Q8G5G9 bifidobacte	620	7	1.7	893	2	O8GCW9	Q8GCW9 chromobacte
548	7	1.7	605	11	O8BH65	O8BH65 mus musculu	621	7	1.7	895	2	O52507	Q52507 serratia 1i
549	7	1.7	610	2	O9S5X1	O9S5X1 streptomyc	622	7	1.7	900	5	O9V7S2	Q9V7S2 drosophila
550	7	1.7	610	2	O69311	O69311 streptomyc	623	7	1.7	901	17	Q9HRV9	Q9HRV9 halobacteri
551	7	1.7	610	16	O86826	O86826 streptomyc	624	7	1.7	917	3	Q9HGS0	Q9HGS0 botrytis ci
552	7	1.7	615	4	O8FFP8	O8FFP8 escherichia	625	7	1.7	938	16	Q8Y1A4	Q8Y1A4 ralstonia s
553	7	1.7	615	4	Q9H7G2	Q9H7G2 homo sapien	626	7	1.7	941	16	O9Z1W5	Q9Z1W5 streptomyc
554	7	1.7	615	4	Q96CM8	Q96CM8 homo sapien	627	7	1.7	947	5	O20625	Q20625 caenorhabdi
555	7	1.7	615	11	O8VCW8	Q8VCW8 mus musculu	628	7	1.7	955	16	O8XW19	Q8XW19 ralstonia s
556	7	1.7	617	16	Q92NXX	Q92NXX rhizobium m	629	7	1.7	959	16	Q8P524	Q8P524 xanthomonas
557	7	1.7	618	11	Q8C5Q0	Q8C5Q0 mus musculu	630	7	1.7	965	5	O9BHC2	Q9BHC2 leishmania
558	7	1.7	619	2	O32374	O32374 clostridium	631	7	1.7	970	12	O8Q206	Q8Q206 rana tigrin
559	7	1.7	620	10	O8RMR6	O8RMR6 triticum ae	632	7	1.7	970	16	O8G6Y0	Q8G6Y0 bifidobacte
560	7	1.7	621	10	Q9MAY8	Q9MAY8 hordeum vul	633	7	1.7	978	16	O8PHB6	O8PHB6 xanthomonas
561	7	1.7	623	10	Q9FVQ2	Q9FVQ2 arabidopsis	634	7	1.7	978	16	O8PE21	O8PE21 xanthomonas
562	7	1.7	624	16	O91237	Q91237 pseudomonas	635	7	1.7	980	10	O8VWK8	O8VWK8 oryza sativ
563	7	1.7	624	16	O8P4H2	Q8P4H2 xanthomonas	636	7	1.7	981	16	Q8XS09	Q8XS09 ralstonia s
564	7	1.7	625	16	O8PG33	Q8PG33 xanthomonas	637	7	1.7	987	10	Q94DRO	Q94DRO oryza sativ
565	7	1.7	626	5	O9NDD1	O9NDD1 leishmania	638	7	1.7	993	11	O8CC22	O8CC22 mus musculu
566	7	1.7	626	16	O8XW42	O8XW42 ralstonia s	639	7	1.7	995	4	O81YU1	O81YU1 homo sapien
567	7	1.7	627	10	Q9M995	Q9M995 arabidopsis	640	7	1.7	997	16	Q98G21	Q98G21 rhizobium 1
568	7	1.7	631	16	O86724	O86724 streptomyc	641	7	1.7	1004	10	O8H384	O8H384 oryza sativ
569	7	1.7	635	2	O30348	O30348 ralstonia s	642	7	1.7	1005	16	Q9HTE6	Q9HTE6 pseudomonas
570	7	1.7	640	5	Q24256	Q24256 drosophila	643	7	1.7	1008	11	Q9DBV8	Q9DBV8 mus musculu
571	7	1.7	645	5	Q9VX55	Q9VX55 drosophila	644	7	1.7	1018	16	Q8EDB9	Q8EDB9 shewanella
572	7	1.7	649	3	O43069	O43069 schizosacch	645	7	1.7	1030	5	O45541	O45541 caenorhabdi
573	7	1.7	658	16	O8BE51	O8BE51 shewanella	646	7	1.7	1041	4	Q96P70	Q96P70 homo sapien
574	7	1.7	660	10	O941T0	Q941T0 oryza sativ	647	7	1.7	1041	16	Q8PHD3	Q8PHD3 xanthomonas
575	7	1.7	662	5	Q9V9F7	Q9V9F7 drosophila	648	7	1.7	1041	16	O8PE13	O8PE13 xanthomonas
576	7	1.7	662	16	Q93RX7	Q93RX7 streptomyc	649	7	1.7	1044	4	O961W3	O961W3 homo sapien
577	7	1.7	668	5	O9BLR2	O9BLR2 leishmania	650	7	1.7	1046	10	O8LHF8	O8LHF8 oryza sativ
578	7	1.7	673	3	O74228	O74228 emericella	651	7	1.7	1053	10	O9PNS3	O9PNS3 chlamydomon
579	7	1.7	674	4	Q96SUC	Q96SUC homo sapien	652	7	1.7	1070	16	O53365	Q53365 synechocyst
580	7	1.7	674	4	O8NCG9	O8NCG9 homo sapien	653	7	1.7	1070	16	O8FEQ0	O8FEQ0 leptospira
581	7	1.7	677	10	Q9M2F5	Q9M2F5 arabidopsis	654	7	1.7	1097	3	O13592	O13592 saccharomyc
582	7	1.7	677	16	O8EJN8	O8EJN8 streptococ	655	7	1.7	1097	16	Q8PP99	Q8PP99 xanthomonas
583	7	1.7	681	16	O9KXG5	O9KXG5 streptomyc	656	7	1.7	1099	2	O68840	O68840 bacillus st
584	7	1.7	682	16	O8XVM2	O8XVM2 ralstonia s	657	7	1.7	1130	4	O96R12	O96R12 homo sapien
585	7	1.7	687	2	O60103	O60103 xanthobacte	658	7	1.7	1136	2	Q9X686	Q9X686 salmonella
586	7	1.7	689	4	O81ZV0	O81ZV0 homo sapien	659	7	1.7	1137	2	Q9AGH9	Q9AGH9 pseudomonas
587	7	1.7	689	11	O8CJ60	O8CJ60 mus musculu	660	7	1.7	1157	4	O96JX7	O96JX7 homo sapien
588	7	1.7	689	16	Q92N56	Q92N56 rhizobium m	661	7	1.7	1173	4	O81X21	O81X21 homo sapien
589	7	1.7	691	5	O9VZE7	O9VZE7 drosophila	662	7	1.7	1207	16	O9A797	O9A797 caulobacter
590	7	1.7	701	2	O9KM03	O9KM03 pseudomonas	663	7	1.7	1213	4	O92923	O92923 homo sapien
591	7	1.7	721	16	O8XB65	O8XB65 escherichia	664	7	1.7	1214	4	O8TAQ2	O8TAQ2 homo sapien
592	7	1.7	721	16	O8FA82	O8FA82 escherichia	665	7	1.7	1215	5	O20626	O20626 caenorhabdi
593	7	1.7	725	2	O8KY01	O8KY01 rhodospendo	666	7	1.7	1237	4	O75147	O75147 homo sapien
594	7	1.7	728	16	O8PH89	O8PH89 xanthomonas	667	7	1.7	1257	5	Q9VK28	Q9VK28 drosophila
595	7	1.7	729	12	O9E114	O9E114 melalegriid h	668	7	1.7	1258	17	O8ZVM3	O8ZVM3 pyrobaculum
596	7	1.7	729	12	Q9DP88	Q9DP88 melalegriid h	669	7	1.7	1345	16	O9Z512	O9Z512 mycobacteri
597	7	1.7	735	10	O81954	O81954 chlamydomon	670	7	1.7	1365	2	O8KWD2	O8KWD2 ruergeria sp
598	7	1.7	738	5	O02402	O02402 pinctada fu	671	7	1.7	1371	16	O8XQ42	O8XQ42 ralstonia s
599	7	1.7	745	16	Q8G6V5	Q8G6V5 bifidobacte	672	7	1.7	1388	5	O8WQ36	O8WQ36 leishmania
600	7	1.7	762	10	Q9SVV7	Q9SVV7 arabidopsis	673	7	1.7	1432	11	O8CHD1	O8CHD1 mus musculu

674	1.7	1530	16	Q9KPU4	Q9KPU4 vibrio chol	747	6	1.5	67	16	Q8EPP9	Q8EPP9 shewanella
675	1.7	1605	2	Q9X6M3	Q9X6M3 salmonella	748	6	1.5	69	2	Q54580	Q54580 streptococc
676	1.7	1734	11	Q6C0592	Q6C0592 mus musculus	749	6	1.5	69	10	Q94K36	Q94K36 arabidopsis
677	1.7	1787	16	Q8Y4Z22	Q8Y4Z22 listeria mo	750	6	1.5	69	12	Q91LD9	Q91LD9 white spot
678	1.7	1788	16	Q92F16	Q92F16 listeria in	751	6	1.5	70	9	Q8SCP3	Q8SCP3 pseudomonas
679	1.7	1963	16	Q8EGK2	Q8EGK2 shewanella	752	6	1.5	70	10	Q8LE23	Q8LE23 arabidopsis
680	1.7	2030	5	Q9VWN8	Q9VWN8 drosophila	753	6	1.5	70	16	Q8RE06	Q8RE06 fusobacteri
681	1.7	2055	2	Q85472	Q85472 abiotrophia	754	6	1.5	71	10	Q8GXU7	Q8GXU7 arabidopsis
682	1.7	2139	5	Q9VU09	Q9VU09 drosophila	755	6	1.5	71	16	Q96UE1	Q96UE1 rhizobium 1
683	1.7	2229	16	Q9E8P5	Q9E8P5 streptomyce	756	6	1.5	71	16	Q97HS9	Q97HS9 clostridium
684	1.7	2365	2	Q93MH9	Q93MH9 streptomyce	757	6	1.5	72	4	Q81WZ9	Q81WZ9 homo sapien
685	1.7	2656	5	Q9GNUM3	Q9GNUM3 paracentrot	758	6	1.5	72	10	Q8H678	Q8H678 oryza sativ
686	1.7	2736	5	Q9NKK6	Q9NKK6 leishmania	759	6	1.5	72	16	P72995	P72995 synechocyst
687	1.7	2946	10	Q64634	Q64634 arabidopsis	760	6	1.5	73	12	Q39543	Q39543 hepatitis g
688	1.7	2957	5	Q61845	Q61845 caenorhabdi	761	6	1.5	73	16	Q9PME7	Q9PME7 campylobact
689	1.7	2993	16	Q8NMS0	Q8NMS0 corynebacte	762	6	1.5	73	16	Q9P905	Q9P905 xyella fas
690	1.7	2998	5	Q8MXK6	Q8MXK6 leishmania	763	6	1.5	73	16	Q92M08	Q92M08 rhizobium m
691	1.7	2998	5	Q9NKS4	Q9NKS4 leishmania	764	6	1.5	73	16	Q8R5T5	Q8R5T5 thermoaerob
692	1.7	3084	12	Q8U211	Q8U211 pseudorabie	765	6	1.5	73	16	Q8P0S1	Q8P0S1 streptococc
693	1.7	3124	4	Q96L91	Q96L91 homo sapien	766	6	1.5	74	2	Q50168	Q50168 mycobacteri
694	1.7	3247	12	Q65553	Q65553 bovine herp	767	6	1.5	75	2	Q8L1D2	Q8L1D2 streptomyce
695	1.7	3436	12	Q66666	Q66666 equine herp	768	6	1.5	75	16	Q8CTP3	Q8CTP3 streptococc
696	1.7	3535	16	Q91791	Q91791 pseudomonas	769	6	1.5	76	16	Q98110	Q98110 rhizobium 1
697	1.7	3591	2	Q9XIE1	Q9XIE1 streptomyce	770	6	1.5	77	9	Q8SBU5	Q8SBU5 bacterioph
698	1.7	3624	16	Q8Z411	Q8Z411 salmonella	771	6	1.5	77	16	Q9PLX2	Q9PLX2 campylobact
699	1.7	3982	2	Q93156	Q93156 bacillus su	772	6	1.5	77	16	Q8FYU0	Q8FYU0 brucella su
700	1.7	4124	5	Q97218	Q97218 leishmania	773	6	1.5	77	17	Q8PYL1	Q8PYL1 methanosa
701	1.7	5627	16	Q91120	Q91120 pseudomonas	774	6	1.5	78	16	Q9CLJ2	Q9CLJ2 pasteurella
702	1.7	9477	2	Q914X3	Q914X3 streptomyce	775	6	1.5	78	16	Q98M09	Q98M09 rhizobium 1
703	1.7	10917	2	Q93NM6	Q93NM6 streptomyce	776	6	1.5	79	16	Q9RKO6	Q9RKO6 streptomyce
704	1.5	31	2	Q9JMV2	Q9JMV2 escherichia	777	6	1.5	80	4	Q8TF08	Q8TF08 homo sapien
705	1.5	34	16	Q8KXV8	Q8KXV8 chlorobium	778	6	1.5	80	5	Q8ISR3	Q8ISR3 spodoptera
706	1.5	35	2	Q8KXN5	Q8KXN5 bacillus an	779	6	1.5	80	16	P71992	P71992 mycobacteri
707	1.5	35	16	Q8VJY6	Q8VJY6 mycobacteri	780	6	1.5	81	5	Q77134	Q77134 drosophila
708	1.5	35	16	Q9FBM7	Q9FBM7 streptomyce	781	6	1.5	81	16	Q9KT78	Q9KT78 vibrio chol
709	1.5	36	6	Q9ETP8	Q9ETP8 leontopithe	782	6	1.5	81	16	Q8X807	Q8X807 escherichia
710	1.5	37	2	Q9EY28	Q9EY28 photobacter	783	6	1.5	81	16	Q8D897	Q8D897 vibrio vuln
711	1.5	40	16	Q8E168	Q8E168 shewanella	784	6	1.5	82	2	Q99Q71	Q99Q71 vibrio harv
712	1.5	42	2	Q9R730	Q9R730 xanthomonas	785	6	1.5	82	2	Q9AJG2	Q9AJG2 vibrio camp
713	1.5	42	10	Q9SQ83	Q9SQ83 arabidopsis	786	6	1.5	82	2	Q54721	Q54721 spirulina p
714	1.5	43	16	Q8XW38	Q8XW38 ralistonia s	787	6	1.5	82	5	Q96486	Q96486 plasmodium
715	1.5	49	6	Q28871	Q28871 canis fam11	788	6	1.5	82	9	Q38378	Q38378 bacterioph
716	1.5	50	16	Q8P0L0	Q8P0L0 streptococc	789	6	1.5	82	9	Q8SBK0	Q8SBK0 bacterioph
717	1.5	50	16	Q8G0A7	Q8G0A7 brucella su	790	6	1.5	83	2	Q85410	Q85410 pseudomonas
718	1.5	51	1	Q9C4W2	Q9C4W2 sulfolobus	791	6	1.5	83	2	Q07490	Q07490 yersinia en
719	1.5	51	4	Q95505	Q95505 homo sapien	792	6	1.5	83	2	Q85417	Q85417 pseudomonas
720	1.5	51	16	Q98G54	Q98G54 rhizobium 1	793	6	1.5	83	10	Q80697	Q80697 arabidopsis
721	1.5	54	4	Q9H2D8	Q9H2D8 homo sapien	794	6	1.5	83	16	Q9A2W0	Q9A2W0 caulobacter
722	1.5	55	6	Q8S086	Q8S086 salmimiri sci	795	6	1.5	83	16	Q8ZCQ0	Q8ZCQ0 salmonella
723	1.5	55	6	Q8S0B4	Q8S0B4 ateles fusc	796	6	1.5	83	16	Q8Z886	Q8Z886 salmonella
724	1.5	55	6	Q8S0H2	Q8S0H2 ateles belz	797	6	1.5	83	16	Q69938	Q69938 streptomyce
725	1.5	55	6	Q8S0Q8	Q8S0Q8 leontopithe	798	6	1.5	85	5	Q9BHC3	Q9BHC3 leishmania
726	1.5	55	6	Q8S0S5	Q8S0S5 lemur varie	799	6	1.5	86	2	Q9S4M5	Q9S4M5 escherichia
727	1.5	55	16	Q9FFY6	Q9FFY6 xyella fas	800	6	1.5	86	9	Q9MCR8	Q9MCR8 bacterioph
728	1.5	56	4	Q16750	Q16750 homo sapien	801	6	1.5	86	16	Q92PD9	Q92PD9 rhizobium m
729	1.5	59	2	Q8VTH2	Q8VTH2 helicobacte	802	6	1.5	87	16	Q8PBL7	Q8PBL7 xanthomonas
730	1.5	59	16	Q8ZRS8	Q8ZRS8 salmonella	803	6	1.5	87	16	Q8NUC8	Q8NUC8 corynebacte
731	1.5	59	16	Q9ZKL5	Q9ZKL5 helicobacte	804	6	1.5	87	16	Q8RUK5	Q8RUK5 corynebacte
732	1.5	60	16	Q8NZT7	Q8NZT7 streptococc	805	6	1.5	88	2	Q56971	Q56971 yersinia pe
733	1.5	60	16	Q8K6X1	Q8K6X1 streptococc	806	6	1.5	88	16	Q8FPU4	Q8FPU4 escherichia
734	1.5	61	12	Q65583	Q65583 bovine herp	807	6	1.5	88	2	Q33552	Q33552 rhodobacter
735	1.5	61	14	Q99156	Q99156 uncultured	808	6	1.5	89	7	Q95956	Q95956 homo sapien
736	1.5	61	16	Q8XNV2	Q8XNV2 clostridium	809	6	1.5	89	7	Q77959	Q77959 homo sapien
737	1.5	61	16	Q8Y106	Q8Y106 brucella me	810	6	1.5	90	9	Q9MCC7	Q9MCC7 bacterioph
738	1.5	62	16	Q8YLD8	Q8YLD8 anabaena sp	811	6	1.5	91	2	Q48265	Q48265 haemophilus
739	1.5	62	16	Q8K799	Q8K799 streptococc	812	6	1.5	91	2	Q85240	Q85240 yersinia en
740	1.5	63	16	Q8BAG4	Q8BAG4 shewanella	813	6	1.5	91	10	Q9SDT4	Q9SDT4 brassica ca
741	1.5	63	16	Q8RFJ1	Q8RFJ1 escherichia	814	6	1.5	91	13	Q8QGS6	Q8QGS6 gallus galli
742	1.5	63	17	Q9HZZ1	Q9HZZ1 halobacteri	815	6	1.5	91	16	Q99QC2	Q99QC2 streptomyce
743	1.5	64	16	Q8RRJ2	Q8RRJ2 fusobacteri	816	6	1.5	91	16	Q8EDT2	Q8EDT2 shewanella
744	1.5	65	16	Q92ZE1	Q92ZE1 rhizobium m	817	6	1.5	92	3	Q59914	Q59914 neotrypodiu
745	1.5	66	16	Q8NTK3	Q8NTK3 corynebacte	818	6	1.5	92	3	Q59910	Q59910 balantia sp
746	1.5	67	4	Q8NB92	Q8NB92 homo sapien	819	6	1.5	92	3	Q59913	Q59913 neotrypodiu

820	6	1.5	92	16	P95120	833	6	1.5	102	16	050482	050482 streptomyc
821	6	1.5	92	16	Q8XBH5	834	6	1.5	103	5	009209	009209 agnus conv
822	6	1.5	92	16	Q9X8E1	835	6	1.5	103	6	09XST6	09XST6 caris faml
823	6	1.5	92	16	Q8ETU6	836	6	1.5	103	16	08XX10	08XX10 ralsconia s
824	6	1.5	93	2	Q92915	837	6	1.5	103	16	0920P9	0920P9 rhizobium m
825	6	1.5	93	16	Q92H17	838	6	1.5	104	16	091RN7	091RN7 arabidopsis
826	6	1.5	94	10	Q9SDU0	839	6	1.5	104	10	08GSP8	08GSP8 oryza sativ
827	6	1.5	94	16	Q92KE9	890	6	1.5	104	16	P73134	P73134 synechocyst
828	6	1.5	95	2	Q48906	901	6	1.5	104	16	09HTW3	09HTW3 pseudomonas
829	6	1.5	95	2	Q50749	902	6	1.5	104	16	098NA6	098NA6 rhizobium l
830	6	1.5	95	2	Q57364	903	6	1.5	104	16	0983R7	0983R7 rhizobium l
831	6	1.5	95	10	Q9SP96	904	6	1.5	105	5	Q25305	Q25305 leishmania
832	6	1.5	95	16	Q93R20	905	6	1.5	105	5	09TVI8	09TVI8 leishmania
833	6	1.5	95	16	Q8DL04	906	6	1.5	105	16	09A4R3	09A4R3 caulobacter
834	6	1.5	96	2	Q9S552	907	6	1.5	105	16	08XVC1	08XVC1 ralsconia s
835	6	1.5	96	2	Q9K534	908	6	1.5	105	16	08FAH2	08FAH2 leptospira
836	6	1.5	96	5	Q9BMV8	909	6	1.5	105	16	08EB76	08EB76 shewanella
837	6	1.5	96	16	Q9CDE1	910	6	1.5	105	17	Q9YE44	Q9YE44 aeropyrum p
838	6	1.5	97	12	Q91HD4	911	6	1.5	106	6	08HY29	08HY29 ovys aries
839	6	1.5	97	12	Q91HD4	912	6	1.5	106	15	041402	041402 human immun
840	6	1.5	97	12	Q91HF4	913	6	1.5	106	16	091630	091630 pseudomonas
841	6	1.5	97	12	Q91HFS	914	6	1.5	106	17	08Z739	08Z739 pyrobaculum
842	6	1.5	97	12	Q91HE4	915	6	1.5	107	5	08SRP1	08SRP1 encephalito
843	6	1.5	97	12	Q91HE2	916	6	1.5	107	16	Q9UYR2	Q9UYR2 neisseria m
844	6	1.5	97	12	Q91HE1	917	6	1.5	107	16	08YXE0	08YXE0 anabaena sp
845	6	1.5	97	12	Q91HG1	918	6	1.5	107	16	08YI88	08YI88 bruceella me
846	6	1.5	97	12	Q91HD7	919	6	1.5	107	16	08YI88	08YI88 ralsconia s
847	6	1.5	97	12	Q91HES	920	6	1.5	107	16	08FZM8	08FZM8 bruceella su
848	6	1.5	97	12	Q91HF7	921	6	1.5	108	3	Q94724	Q94724 schizosacch
849	6	1.5	97	12	Q91HF3	922	6	1.5	108	5	Q26290	Q26290 aedes aegy
850	6	1.5	97	12	Q91HF1	923	6	1.5	108	13	08AYB4	08AYB4 salvelinus
851	6	1.5	97	12	Q91HD8	924	6	1.5	108	16	092W48	092W48 rhizobium m
852	6	1.5	97	12	Q91HE9	925	6	1.5	108	17	09Y922	09Y922 aeropyrum p
853	6	1.5	97	12	Q91HE0	926	6	1.5	109	2	Q9AJH0	Q9AJH0 vibrio prot
854	6	1.5	97	12	Q91HE8	927	6	1.5	109	2	Q9ANF1	Q9ANF1 bradyrhizob
855	6	1.5	97	12	Q91HE7	928	6	1.5	109	6	Q29286	Q29286 sus scrofa
856	6	1.5	97	12	Q91HD3	929	6	1.5	109	9	Q94MQ9	Q94MQ9 bacterioph
857	6	1.5	97	12	Q91HD6	930	6	1.5	109	16	Q25421	Q25421 helicobacte
858	6	1.5	97	12	Q91HP9	931	6	1.5	109	16	08NMV2	08NMV2 cornebacte
859	6	1.5	97	12	Q91HP8	932	6	1.5	109	16	Q9X948	Q9X948 streptomyc
860	6	1.5	97	12	Q91HE3	933	6	1.5	110	2	Q54878	Q54878 streptococ
861	6	1.5	97	12	Q91HE6	934	6	1.5	110	2	Q33004	Q33004 mycobacteri
862	6	1.5	97	12	Q91HD5	935	6	1.5	110	5	08ISM2	08ISM2 myrmecia ni
863	6	1.5	97	12	Q91HD5	936	6	1.5	110	12	08OW60	08OW60 norwayk-lik
864	6	1.5	97	12	Q91HE2	937	6	1.5	110	16	08YOE4	08YOE4 ralsconia s
865	6	1.5	97	12	Q91HD9	938	6	1.5	110	16	08PER0	08PER0 xanthomonas
866	6	1.5	97	12	Q91HE6	939	6	1.5	110	17	08ZSN0	08ZSN0 pyrobaculum
867	6	1.5	98	12	P89937	940	6	1.5	111	10	Q94ZU5	Q94ZU5 oryza sativ
868	6	1.5	98	12	P90501	941	6	1.5	112	2	Q91BE9	Q91BE9 helicobacte
869	6	1.5	98	16	Q92T89	942	6	1.5	112	10	08L4D1	08L4D1 oryza sativ
870	6	1.5	98	16	Q8RD90	943	6	1.5	112	11	09QY78	09QY78 raltus norv
871	6	1.5	98	16	Q8F3E5	944	6	1.5	112	16	Q92K70	Q92K70 rhizobium m
872	6	1.5	99	16	Q8F3E5	945	6	1.5	112	16	006774	006774 mycobacteri
873	6	1.5	99	13	Q8AYJ2	946	6	1.5	113	2	Q9X9U7	Q9X9U7 streptomyc
874	6	1.5	99	13	Q8AYJ2	947	6	1.5	113	16	Q98XK14	Q98XK14 rhizobium l
875	6	1.5	99	16	Q8ECT8	948	6	1.5	113	16	08XCT2	08XCT2 escherichia
876	6	1.5	100	2	Q50146	949	6	1.5	113	16	08ZSM0	08ZSM0 yersinia pe
877	6	1.5	100	2	Q9FP94	950	6	1.5	113	16	08CVZ6	08CVZ6 escherichia
878	6	1.5	100	8	Q32758	951	6	1.5	114	2	Q8VL88	Q8VL88 bacillus su
879	6	1.5	100	10	Q9FUZ8	952	6	1.5	114	2	08VOA5	08VOA5 bacillus su
880	6	1.5	100	12	Q9WB51	953	6	1.5	114	2	08VOA7	08VOA7 bacillus su
881	6	1.5	100	12	Q9DUC2	954	6	1.5	114	4	08VOA4	08VOA4 bacillus su
882	6	1.5	101	5	Q9NKO9	955	6	1.5	114	8	Q9MFE4	Q9MFE4 beta vulgar
883	6	1.5	101	6	Q9XSO2	956	6	1.5	114	11	Q91YF3	Q91YF3 mus musculu
884	6	1.5	101	16	Q9CDX3	957	6	1.5	114	12	011315	011315 moliniscum c
885	6	1.5	101	16	Q8EIM9	958	6	1.5	114	13	Q9DFP8	Q9DFP8 oncorhynch
886	6	1.5	101	16	Q8E7T0	959	6	1.5	115	2	Q51045	Q51045 neisseria g
887	6	1.5	101	16	Q8E5G8	960	6	1.5	115	9	Q9AZA5	Q9AZA5 rhodococcu
888	6	1.5	101	16	Q8E2C3	961	6	1.5	115	9	Q9AZA5	Q9AZA5 lactobacilli
889	6	1.5	101	16	Q8DZR9	962	6	1.5	115	10	08Z787	08Z787 oryza sativ
890	6	1.5	101	16	Q8DS24	963	6	1.5	115	10	Q96NO9	Q96NO9 oryza sativ
891	6	1.5	102	10	Q8S4O9	964	6	1.5	115	16	Q9RXJ1	Q9RXJ1 deinococcus
892	6	1.5	102	11	Q99XG2	965	6	1.5	115	16	Q8P4B2	Q8P4B2 xanthomonas

966	6	1.5	115	16	Q91210	091210 streptomyc
967	6	1.5	116	2	Q51016	051016 neisseria g
968	6	1.5	116	2	Q8GB96	Q8GB96 heliobacill
969	6	1.5	116	10	Q9SNR0	Q9SNR0 oryza sativ
970	6	1.5	116	10	Q9SNR7	Q9SNR7 oryza sativ
971	6	1.5	116	10	Q92TR3	Q92TR3 physalis ci
972	6	1.5	116	10	Q8GBZ0	Q8GBZ0 triticum ae
973	6	1.5	116	16	Q9RV55	Q9RV55 deinococcus
974	6	1.5	116	16	Q9X9T5	Q9X9T5 streptomyce
975	6	1.5	116	17	Q9H0G9	Q9H0G9 halobacteri
976	6	1.5	117	2	Q51017	Q51017 neisseria g
977	6	1.5	117	2	Q9WTG4	Q9WTG4 escherichia
978	6	1.5	117	16	Q92WZ6	Q92WZ6 rhizobium m
979	6	1.5	117	16	Q8PFW6	Q8PFW6 xanthomonas
980	6	1.5	117	16	Q9RDE3	Q9RDE3 streptomyce
981	6	1.5	118	2	Q93H11	Q93H11 streptomyce
982	6	1.5	118	10	Q8S745	Q8S745 oryza sativ
983	6	1.5	118	12	Q11279	Q11279 human papil
984	6	1.5	118	16	Q9A7X0	Q9A7X0 caulobacter
985	6	1.5	118	16	Q8Y1U5	Q8Y1U5 ralslonia s
986	6	1.5	119	2	Q9AN60	Q9AN60 bradyrhizob
987	6	1.5	119	2	Q52505	Q52505 pseudomonas
988	6	1.5	119	16	Q9PFZ4	Q9PFZ4 xyella fas
989	6	1.5	119	16	Q8ZKK2	Q8ZKK2 salmonella
990	6	1.5	119	16	Q97GR3	Q97GR3 clostridium
991	6	1.5	119	16	Q8DJ95	Q8DJ95 synechococc
992	6	1.5	120	2	P74898	P74898 thermus the
993	6	1.5	120	10	Q8LHY6	Q8LHY6 oryza sativ
994	6	1.5	121	3	Q94217	Q94217 cladosporiu
995	6	1.5	121	8	Q9TG24	Q9TG24 rhinopolus
996	6	1.5	121	9	Q80111	Q80111 bacteriophu
997	6	1.5	121	10	Q8GW10	Q8GW10 arabidopsis
998	6	1.5	121	16	Q9KAG1	Q9KAG1 bacillus ha
999	6	1.5	121	16	Q8Y357	Q8Y357 ralslonia s
1000	6	1.5	121	17	Q96ZP1	Q96ZP1 sulfobobus

ALIGNMENTS

RESULT 1

Q9JTS0 PRELIMINARY; PRT; 412 AA.
AC Q9JTS0;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Membrane fusion protein.
GN MTRC OR NMA1970.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OK NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=22491 / Serogroup A / Serotype 4A;
RX MEDLINE=2022556; PubMed=10761919;
RA Parthill J., Achman M., James K.D., Bentley S.D., Churcher C., Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T., Davids R.M., Davis P., Devlin K., Feltham T., Hamlin N., Holtroyd S., Jags R., Leather S., Moul S., Mungall K., Quail M.A., Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J., Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491."
RL Nature 404:502-506(2000).
DR EMBL; AL162757; CAB85190.1; -
DR InterPro; IPR006143; H1YD.
DR Pfam; PF00529; H1YD; 1.
KW Complete proteome.
SQ SEQUENCE 412 AA; 42954 MW; 5CF797BA370AA75D CRC64;

Query Match 41.5%; Score 171; DB 16; Length 412;

Best Local Similarity 99.6%; Pred. No. 6,66-151;									
Matches 271; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
QY	1	MAFAFRKRAAALAAVAATLVS	CGKGGDAAOGGPRAPRVGVVTHPQTVALT	60					
Db	1	MAFAFRKRAAALAAVAATLVS	CGKGGDAAOGGPRAPRVGVVTHPQTVALT	60					
QY	61	VELGRLESLRTADVRAOVGGII	IQKRLFQESSYVRACGPLYQIISSTYEANLIESAPQOLA	120					
Db	61	VELGRLESLRTADVRAOVGGII	IQKRLFQESSYVRACGPLYQIISSTYEANLIESAPQOLA	120					
QY	121	TAQATLAKADADLARVYKPLVA	AEAVSRQEDAAVTAKRSAPGVKAAQAIAKAGINLR	180					
Db	121	TAQATLAKADADLARVYKPLVA	AEAVSRQEDAAVTAKRSAPGVKAAQAIAKAGINLR	180					
QY	181	SRITAPISGFIGQSKVSEGTLL	NAGDTTLATITQTPMNVNQSASEWKLKRLQIAEG	240					
Db	181	SRITAPISGFIGQSKVSEGTLL	NAGDTTLATITQTPMNVNQSASEWKLKRLQIAEG	240					
QY	241	KLLAADGVIAVGIRFDDGVY	PEKGRLLFADP 272						
Db	241	KLLAADGVIAVGIRFDDGVY	PEKGRLLFADP 272						

RESULT 2

Q9JY66 PRELIMINARY; PRT; 412 AA.
AC Q9JY66;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Membrane fusion protein.
GN NMB1716.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OK NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tetzelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E., Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J., Nelson W.C., Gwin M.L., DeBoy R., Peterson J.D., Hickey E.R., Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A., Mason T., Ciecko A., Parksey D.S., Blair E., Citterone H., Clark E.B., Cotton M.D., Ullrichback T.R., Knouri H., Qin H., Yamathavan J., Gill J., Scarlato V., Maignani V., Pizsa M., Grandi G., Sun L., Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain MC58."
RL Science 287:1809-1815(2000).
DR EMBL; AE002521; AAF42063.1; -
DR TIGR; NMB1716; -
DR InterPro; IPR006143; H1YD.
DR Pfam; PF00529; H1YD; 1.
KW Complete proteome.
SQ SEQUENCE 412 AA; 42795 MW; 087A0DB0E3113E4E CRC64;

Query Match 32.3%; Score 133; DB 16; Length 412;
Best Local Similarity 100.0%; Pred. No. 2e-115;
Matches 133; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	274	VNESTGQITLRAAVNDONITL	MPGLVYVLMDOVAVDNAFVVPQOAVTRGAKDTVMIVNA	333					
Db	274	VNESTGQITLRAAVNDONITL	MPGLVYVLMDOVAVDNAFVVPQOAVTRGAKDTVMIVNA	333					
QY	334	OGMEPREVTAAGQGTMI	VTSGLKDGDKVYVGGISITAGITGAKVTPKEMASSENOAA	393					
Db	334	OGMEPREVTAAGQGTMI	VTSGLKDGDKVYVGGISITAGITGAKVTPKEMASSENOAA	393					
QY	394	APQSGVGTASEAK	406						

Db 394 APOGCVQTASEAK 406

RESULT 3

OS1007 PRELIMINARY; PRT; 271 AA.

AC 051007; PRELIMINARY; PRT; 271 AA.

DT 01-NOV-1996 (TRENBLrel. 01, Created)

DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)

DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)

DE Mtrc protein (Fragment).

GN MTRC.

OS Neisseria gonorrhoeae.

OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;

OC Neisseriaceae; Neisseria.

OX NCBI_TaxID=485;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CH95;

RX MEDLINE=94254732; PubMed=8196548;

RA Pan W., Spratt B.G.;

RT "Regulation of the permeability of the gonococcal cell envelope by the mtr system.";

RL Mol. Microbiol. 11:769-775(1994).

DR EMBL; 225796; CA81046.1; -.

DR InterPro; IPR006143; H1YD.

DR Pfam; PF00529; H1YD; 1.

FT NON TER 271

FT SEQUENCE 271 AA; 28145 MW; 1DDADD1F335B39C5 CRC64;

Query Match 24.3%; Score 100; DB 2; Length 271;

Best Local Similarity 100.0%; Pred. No. 9.1e-85;

Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 OEYDAVAVAKSAEAGVKAQAAIKSAGININRSRTTAPISGFIQSKVSEGTLLNAGDT 207

Db 148 OEYDAVAVAKSAEAGVKAQAAIKSAGININRSRTTAPISGFIQSKVSEGTLLNAGDT 207

QY 208 TVLATIRGTNPYVAVTOSASEVMTLRQIAEGKLIADG 247

Db 208 TVLATIRGTNPYVAVTOSASEVMTLRQIAEGKLIADG 247

RESULT 4

Q8X7E1 PRELIMINARY; PRT; 373 AA.

AC 08X7E1; PRELIMINARY; PRT; 373 AA.

DT 01-MAR-2002 (TRENBLrel. 20, Created)

DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)

DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)

DE Putative efflux pump (Putative multidrug-efflux transport protein).

GN Z2509 OR EGS1863.

OS Escherichia coli O157:H7.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

OX NCBI_TaxID=83334;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / EDL933 / ATCC 700927;

RX MEDLINE=21074935; PubMed=11206551;

RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D., Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A., Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L., Groetbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamocous K., Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C., Welch R.A., Blattner F.R.;

RA "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";

RT Nature 409:529-533(2001).

RL [2]

RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / RIMD 0509952;

RX MEDLINE=21156231; PubMed=11258796;

RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K., Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T., RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T., RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;

RT "Complete genome sequence of enterohaemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12.";

RL DNA Res. 8:11-22(2001).

DR EMBL; AE005377; AAG56521.1; -.

DR EMBL; AP002556; BAB35286.1; -.

DR InterPro; IPR006143; H1YD.

DR Pfam; PF00529; H1YD; 1.

KM Complete proteome.

SQ SEQUENCE 373 AA; 39612 MW; 2A573537AB44CDBF CRC64;

Query Match 2.7%; Score 11; DB 16; Length 373;

Best Local Similarity 100.0%; Pred. No. 0.16;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 78 QVGGIIQKRLP 88

Db 63 QVGGIIQKRLP 73

RESULT 5

Q8GC84 PRELIMINARY; PRT; 374 AA.

AC 08GC84; PRELIMINARY; PRT; 374 AA.

DT 01-MAR-2003 (TRENBLrel. 23, Created)

DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)

DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)

DE Befa lipoprotein precursor.

GN BEFA.

OS Enterobacter aerogenes (Aerobacter aerogenes).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Enterobacter.

OX NCBI_TaxID=548;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BW16627;

RA Maai M., Pages J.M., Pradel E.;

RT "Identification and characterization of the Enterobacter aerogenes RT eefABC operon encoding a putative tripartite efflux system.";

RT Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.

DR EMBL; AJ508047; CAD4861.1; -.

KW Signal; Lipoprotein.

FT SIGNAL 1 17

FT CHAIN 18 374

FT SEQUENCE 374 AA; 39374 MW; 332BFD6FAB942BB CRC64;

Query Match 2.7%; Score 11; DB 2; Length 374;

Best Local Similarity 100.0%; Pred. No. 0.16;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 78 QVGGIIQKRLP 88

Db 64 QVGGIIQKRLP 74

RESULT 6

Q8CM42 PRELIMINARY; PRT; 384 AA.

AC 08CM42; PRELIMINARY; PRT; 384 AA.

DT 01-MAR-2003 (TRENBLrel. 23, Created)

DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)

DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)

DE Acriflavine resistance protein A precursor.

GN C1761.

OS Escherichia coli O6.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

OX NCBI_TaxID=217992;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=06.H1 / CFT073 / ATCC 700928;
 RX MEDLINE=2388234; PubMed=12471157;
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
 RA Rasko D., Buckles E.L., Liu S.-R., Boutin A.C., Hackett J., Stroud D.,
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 of uropathogenic *Escherichia coli*.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
 KW EMBL; AF016760; AAN80227.1; -.
 SQ SEQUENCE 384 AA; 41129 MW; FC9E69E4583EC7BB CRC64;

Query Match 2.7%; Score 11; DB 16; Length 384;
 Best Local Similarity 100.0%; Pred. No. 0.16;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 78 QVGGIIQKRLF 88
 |||||
 DB 74 QVGGIIQKRLF 84

RESULT 7
 ID Q9AA03 PRELIMINARY; PRT; 388 AA.

AC Q9AA03; 01-JUN-2001 (TREMblrel. 17, Created)
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
 DT 01-OCT-2002 (TREMblrel. 22, Last annotation update)
 DE HlyD family secretion protein.
 GN CC0808.

OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
 CC Caulobacteriaceae; Caulobacter.
 OX NCBI_TaxID=155892;

(1)
 RP SEQUENCE FROM N.A.
 RN RC STRAIN=ATCC 19089 / CB15;
 RX MEDLINE=21173698; PubMed=11259647;
 RA Nieman W.C., Fejdylyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
 RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
 RA Deboy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
 RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
 RA Ueberback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
 RT "Complete genome sequence of *Caulobacter crescentus*.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141 (2001).
 DR EMBL; AB005757; AAK22793.1; -.
 DR TIGR; CC0808; -.
 DR InterPro; IPR006143; HlyD.
 DR Pfam; PF00529; HlyD; 1.
 KW Complete proteome.
 SQ SEQUENCE 388 AA; 39483 MW; 4D934C3F69052DAD CRC64;

Query Match 2.7%; Score 11; DB 16; Length 388;
 Best Local Similarity 100.0%; Pred. No. 0.16;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 94 VRAAGPLVQID 104
 |||||
 DB 89 VRAAGPLVQID 99

RESULT 8
 ID Q969P9 PRELIMINARY; PRT; 124 AA.

AC Q969P9; 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DT 01-OCT-2002 (TREMblrel. 22, Last annotation update)
 DE CkII beta binding protein 2.
 OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung, and Muscle;
 RA Strausberg R.;
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Yoon S.-H., Bae Y.-S.;
 RT "Homo sapiens CXBBP2 mRNA.";
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC013039; AAH13039.1; -.
 DR EMBL; BC013027; AAH13027.1; -.
 DR EMBL; AF412816; AAL55291.1; -.
 SQ SEQUENCE 124 AA; 14494 MW; 9C50C85B061DFD4A CRC64;

Query Match 2.4%; Score 10; DB 4; Length 124;
 Best Local Similarity 100.0%; Pred. No. 0.51;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 RAALAAVA 19
 |||||
 DB 104 RAALAAVA 113

RESULT 9
 ID Q81VM3 PRELIMINARY; PRT; 216 AA.

AC Q81VM3; 01-MAR-2003 (TREMblrel. 23, Created)
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE Hypothetical protein (fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;

(1)
 RP SEQUENCE FROM N.A.
 RN RA Revolella C.;
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Frigimelica E., Lanfranchi G.;
 RT "Full length sequencing of some human and murine muscular transcripts
 (Telethon Italy Project B41).";
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ295986; CAC82499.1; -.
 KW Hypothetical protein.
 FT NON TER 1
 SQ SEQUENCE 216 AA; 24768 MW; D7DE609D6188B835 CRC64;

Query Match 2.4%; Score 10; DB 4; Length 216;
 Best Local Similarity 100.0%; Pred. No. 0.83;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 RAALAAVA 19
 |||||
 DB 196 RAALAAVA 205

RESULT 10
 ID Q9BSM6 PRELIMINARY; PRT; 218 AA.

AC Q9BSM6; 01-JUN-2001 (TREMblrel. 17, Created)
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE Similar to RIKEN cDNA 2310040G17 gene (fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=skin;
RA Strusberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC004944; AA04944.1; -.
FT NON TER
SQ SEQUENCE 218 AA; 25023 MW; 71B4A2752F4CFEDB CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 0.84;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 RAALAAVA 19
Db 198 RAALAAVA 207

RESULT 11
OBTES1 PRELIMINARY; PRT; 222 AA.
AC OBTES1;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Papillomavirus L2 interacting nuclear protein 1.
CN PLINP-1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Mueller M.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Goernemann J.;
RL "Cellular Interaction Partners of the HPV Minor Capsid Protein L2."
RL Thesis (2002), Department of Fakultät fuer Biologie,
RL Universität Heidelberg, Heidelberg, Germany.
DR EMBL; AJ429498; CAD2344.1; -.
KW Nuclear protein.
SQ SEQUENCE 222 AA; 25368 MW; 21EF67045B1F5868 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 0.85;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 RAALAAVA 19
Db 202 RAALAAVA 211

RESULT 12
OBTES1 PRELIMINARY; PRT; 222 AA.
AC OBTES1;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE CR6 interacting factor 1 (CKI1 beta associating protein).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Chung H., Yi Y., Jung N.-C., Kim D., Shong M.;
RL "CR6 interacting factor 1 (CRIF1), a novel nuclear protein that
interacts with Gadd45 family proteins. Cloning, functional

```

```

RT characterization and analysis of expression."
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Yoon S.H., Bae Y.S.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF479749; AAL85877.1; -.
DR EMBL; AF475095; AAM10639.1; -.
SQ SEQUENCE 222 AA; 25384 MW; 304C397AEC958804 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 0.85;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 RAALAAVA 19
Db 202 RAALAAVA 211

RESULT 13
OBTES1 PRELIMINARY; PRT; 444 AA.
AC OBTES1;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DE 01-MAR-2002 (TREMBlrel. 23, Last annotation update)
DE Putative HlyD family secretion protein (Putative membrane
protein).
CN IPO2847 OR Y1386.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titchall R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Fellwell T., Hamlin N., Holroyd S., Jagsels K., Katlyshev A.V.,
RA Leather S., Moul S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrett B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague."
RL Nature 413:523-527(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=KIMS / Biovar Mediaevalis;
RX MEDLINE=22137863; PubMed=12142430;
RA Deng W., Burland V., Plunkett G., III, Boutin A., Mayhew G.F., Liss P.,
RA Perna N.T., Rose D.V., Mau B., Zhou S., Schwartz D.C.,
RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
RA Perry R.D.;
RT "Genome sequence of Yersinia pestis KIM."
RL J. Bacteriol. 184:4601-4611(2002).
DR EMBL; AJ414154; CAC92099.1; -.
DR EMBL; AE013741; AAM84958.1; -.
DR InterPro; IPR006143; HlyD.
DR Pfam; PF00529; HlyD; 1.
KW Complete proteome.
SQ SEQUENCE 444 AA; 47514 MW; BF70B791FF55BD01 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 13;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 181 SRTAPISG 189
Db 198 SRTAPISG 206

```

```

RESULT 14
ID 09J007 PRELIMINARY: PRT; 107 AA.
AC 09J007;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DE 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE 50S ribosomal protein L24.
GN RPLX OR NMA0118 OR NMB0153.
OS Neisseria meningitidis (serogroup A), and
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=6569, 491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2491 / Serogroup A / Serotype 4A;
RX MEDLINE=20222556; PubMed10761919;
RA Parthill J., Achman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltham T., Hamilton N., Holtroyd S.,
RA Jags R., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajendram M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrett B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis 22491."
RL Nature 404:502-506 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=20175755; PubMed10710307;
RA Tetteijn H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eissen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwin M.L., Debey R., Peterson J.D., Hickey E.K.,
RA Hafe D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Clecko A., Parksey D.S., Blair E., Clifton H., Clark E.B.,
RA Cotton M.D., Utechtack T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scariato V., Masigiani V., Pizsa M., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RL Science 287:1809-1815 (2000).
DR EMBL; AL162752; CAB83433.1; -.
DR EMBL; AE002373; AAF40611.1; -.
DR TIGR; NMB0153; -.
DR InterPro; IPR005824; KOW.
DR InterPro; IPR006646; KOW_sub.
DR InterPro; IPR003256; RIBosomal_L24.
DR InterPro; IPR005825; RIBosomal_L24_26.
DR Pfam; PF00467; KOW; 1.
DR ProDom; PD001677; RIBosomal_L24; 1.
DR SMART; SM00739; KOW; 1.
DR TIGRfams; TIGR01079; rplX_bact; 1.
DR PROSITE; PS01108; RIBOSOMAL_L24; 1.
KW RIBosomal protein; Complete proteome.
SQ SEQUENCE 107 AA; 11667 MW; EICEAA2FOA93A24C CRC64;

Query Match 1.9%; Score 8; DB 16; Length 107;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

DE P0712E02.9 protein (OSUNB0024F06.19 protein).
GN P0712E02.9 OR OSUNB0024F06.19.
OS Oryza sativa (Rice), and
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530, 39947;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=O. sativa; STRAIN=cv. Nipponbare;
RA Saeki T., Matsunoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (Ga3) genomic DNA, chromosome 1, PAC
RT clone:P0712E02."
RL Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=O. sativa (japonica cultivar-group); STRAIN=cv. Nipponbare;
RA Saeki T., Matsunoto T., Yamamoto K.;
RT "Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, BAC
RT clone:OSUNB0024F06."
RL Submitted (Feb-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AP003492; BAB1846.1; -.
DR EMBL; AP003341; BAC01241.1; -.
DR Gramene; Q94D46; -.
SQ SEQUENCE 125 AA; 12949 MW; 0488F2AB81F7C6C0 CRC64;

Query Match 1.9%; Score 8; DB 10; Length 125;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 15 AAAVALVL 22
DB 30 AAAVALVL 37

RESULT 16
ID 032184 PRELIMINARY: PRT; 129 AA.
AC 032184;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DE 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE YUSR protein.
GN YUSR.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunet F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Boloctin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brigelli S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabre C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Haele J., Harwood C.R., Henaut A.,
RA Hilbert H., Holstappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaert-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,

```

```

RA Saitouchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
RA Sorokin A., Taconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Toesato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambuit R., Wedler E., Wedler H., Weitzenecker T.,
RA Winers P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zimstein B., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis".
RL Nature 390:249-256(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
DR EMBL; Z99120; CAB15279.1; -.
DR HSSP; 070351; 186W.
DR InterPro; IPR002198; ADH_short.
DR Pfam; PF00106; adh_short; 1.
DR PRINTS; PR00080; SDRFAMILY.
DR PROSITE; PS00061; ADH_SHORT; 1.
KW Oxidoreductase; Complete proteome.
SQ SEQUENCE 129 AA; 13477 MW; 1874C97295B292A CRC64;

Query Match
Best Local Similarity 1.9%; Score 8; DB 16; Length 129;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 370 SIAGITGA 377
Db 28 SIAGITGA 35

RESULT 17
O9RPUI PRELIMINARY; PRT; 156 AA.
AC O9RPUI;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE Aerobically induced outer membrane protein AipI (Fragment).
GN AipI.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TY2;
RA Maulen N.P., Mora G.C.;
RT "An anaerobically induced outer membrane protein of Salmonella typhi
RT is necessary for the efficient invasion of Hep-2 cells."
RL Submitted (APR-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF145591; AAF08700.1; -.
FT NON_TER 1 156
FT NON_TER 1 156
SQ SEQUENCE 156 AA; 16591 MW; 1EC862A52419ECCB CRC64;

Query Match
Best Local Similarity 1.9%; Score 8; DB 2; Length 156;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 10 RAALAAA 17
Db 13 RAALAAA 20

RESULT 18
O8SOKI PRELIMINARY; PRT; 162 AA.
AC O8SOKI;

```

```

DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE B1078G07.34 protein (P0696G06.11 protein).
GN B1078G07.34 OR P0696G06.11.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, BAC
RT clone:B1078G07."
RL Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
RT clone:P0696G06."
RL Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AP003407; BAB90212.1; -.
DR EMBL; AP003316; BAC06254.1; -.
DR Gramene; Q8SOKI; -.
SQ SEQUENCE 162 AA; 18253 MW; 873B89784D35DBF4 CRC64;

Query Match
Best Local Similarity 1.9%; Score 8; DB 10; Length 162;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 15 AAAVALVL 22
Db 8 AAAVALVL 15

RESULT 19
O8XCR9 PRELIMINARY; PRT; 171 AA.
AC O8XCR9;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Putative fibrial protein (Putative fibrial-like protein).
GN SPMF OR 20691 OR ECS0596.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / R1MD 0509952;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Apodaca J., Davis N.W., Lim A., DiMalanta E.T., Potamousis K.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / R1MD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayaishi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kihara S., Shiba T., Hatori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."

```

RL DNA Res. 8:11-22(2001).
 DR EMBL; AB005234; AAG54891.1; -
 DR EMBL; AP002552; BAB34019.1; -
 DR InterPro: IPR002559; Fimbrilal.
 DR Pfam; PF00419; Fimbrilal; 1.
 KW Complete proteome.
 SQ SEQUENCE 171 AA; 18212 MW; 629C9DC02FC83516 CRC64;

Query Match 1.9%; Score 8; DB 16; Length 171;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 201 LFNAGDTT 208
 |||||
 Db 61 LFNAGDTT 68

RESULT 20

O8XC98 PRELIMINARY; PRT; 183 AA.
 AC O8XC98;
 DT 01-MAR-2002 (TRENBLrel. 20, Created)
 DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
 DE Hypothetical protein z2047.
 GN z2047 OR ECOL1767.
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; Escherichia.
 OK NCBI_TaxID=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;

RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Poštal G., Hackert J., Klink S., Boutin A., Shao Y., Miller B.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
 RL Nature 409:529-533(2001).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
 RA Kohara S., Shiba T., Hattori M., Shingawa H.;
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 RT O157:H7 and genomic comparison with a laboratory strain K-12."
 RL DNA Res. 8:11-22(2001).
 DR EMBL; AB005343; AAG56122.1; -
 DR EMBL; AP002556; BAB35190.1; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 183 AA; 21220 MW; 5A6C3E3B9000796E CRC64;

Query Match 1.9%; Score 8; DB 16; Length 183;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 267 LFNADPVV 274
 |||||
 Db 130 LFNADPVV 137

RESULT 21

O9YAH2 PRELIMINARY; PRT; 183 AA.
 AC O9YAH2;
 DT 01-NOV-1999 (TRENBLrel. 12, Created)
 DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)

DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
 DE Hypothetical protein ABE1967.
 GN ABE1967.
 OS Aeropyrum pernix.
 OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcates;
 CC Desulfurococcaceae; Aeropyrum.
 OK NCBI_TaxID=56636;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K1;
 RX MEDLINE=99331039; PubMed=10382966;
 RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
 RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Anka A., Kosugi H.,
 RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
 RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
 RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
 RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
 RT "Complete genome sequence of an aerobic hyper-thermophilic
 RT crenarchaeon, Aeropyrum pernix K1."
 RL DNA Res. 6:83-101(1999).
 DR EMBL; AP000063; BAA80977.1; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 183 AA; 19724 MW; 60408554DBA0ABC CRC64;

Query Match 1.9%; Score 8; DB 17; Length 183;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 13 ALAAVAL 20
 |||||
 Db 118 ALAAVAL 125

RESULT 22

O8P176 PRELIMINARY; PRT; 206 AA.
 AC O8P176;
 DT 01-OCT-2002 (TRENBLrel. 22, Created)
 DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
 DE Hypothetical protein XAC3022.
 GN XAC3022.
 OS Xanthomonas axonopodis (pv. citri).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 CC Xanthomonadaceae; Xanthomonas.
 OK NCBI_TaxID=92829;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=306 / ATCC 13902 / XV 101;
 RX MEDLINE=22022145; PubMed=12024217;
 RA da Silva A.C.R., Ferro J.A., Reinach P.C., Farah C.S., Furlan L.R.,
 RA Alves L.M.C., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
 RA Camargo L.E.A.;
 RA Camarotte G., Cammaran F., Cardoso J., Chamargo F., Ciapina L.P.,
 RA Ciccarelli R.M.B., Coutinho L.B., Cursino-Santos J.R., El-Dorri H.,
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
 RA Forghieri E.F., Franco M.C., Greggio C.C., Gruber A.,
 RA Katayama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
 RA Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
 RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
 RA Pereira H.A., Rossi A., Sana J.A.D., Silva C., de Souza R.F.,
 RA Spínola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
 RA Trindade dos Santos M., Truffi D., Teal S.M., White F.F.,
 RA Setubal J.C., Kitajima J.P.;
 RT "Comparison of the genomes of two Xanthomonas pathogens with differing
 RT host specificities."
 RL Nature 417:459-463(2002).
 DR EMBL; AB011943; AAM37867.1; -
 DR InterPro: IPR00104; Antifreeze_1.
 DR PRINTS; PR00308; ANTIFREEZE1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 206 AA; 20298 MW; 88C31E39F624A908 CRC64;

Query Match 1.9%; Score 8; DB 16; Length 206;
 Best Local Similarity 100.0%; Pred. No. 59;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 RAALAA 17
 |||||
 Db 32 RAALAA 39

RESULT 23

O9KZV1 PRELIMINARY; PRT; 206 AA.

AC O9KZV1
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
 DE Putative ABC transport system transmembrane protein.
 GN SCO4150 OR SCD84.17.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 CX NCBI_TaxID=1902;
 [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RX MEDLINE=2196410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrall B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 RT coelicolor A3(2).";
 RT Nature 417:141-147(2002).
 DR EMBL: AL939119; CAB88483.1; -.
 KW Transmembrane; Complete proteome.
 SQ SEQUENCE 206 AA; 21043 MW; A3D81PAD92828A39 CRC64;

Query Match 1.9%; Score 8; DB 16; Length 206;
 Best Local Similarity 100.0%; Pred. No. 59;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 14 LAAAVLV 21
 |||||
 Db 184 LAAAVLV 191

RESULT 24

P95115 PRELIMINARY; PRT; 214 AA.

AC P95115
 DT 01-MAY-1997 (TREMblrel. 03, Created)
 DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
 DE Hypothetical protein RV2980.
 GN RV2980 OR MT3058 OR MTCY349.07C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 CX NCBI_TaxID=1773;
 [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigemeier K., Gao S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,

RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sultun J.E., Taylor K., Whitehead S., Barrall B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RT Nature 393:537-544(1998).
 [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., Deboy R., Dodson R., Gwin M., Holt D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains.";
 RT Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Z83018; CAB05432.1; ALT INIT.
 DR EMBL: AE007126; AAK47385.1; -.
 DR TIGR: MT3058; -.
 DR TubercuList; RV2980; -.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 214 AA; 22522 MW; EDD46F75C351D96E CRC64;

Query Match 1.9%; Score 8; DB 16; Length 214;
 Best Local Similarity 100.0%; Pred. No. 61;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 AAALAAV 18
 |||||
 Db 49 AAALAAV 56

RESULT 25

ID O8SPE7 PRELIMINARY; PRT; 222 AA.

AC O8SPE7
 DT 01-JUN-2002 (TREMblrel. 21, Created)
 DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Putative papillomavirus L2 interacting nuclear protein.
 GN PLINP.
 OS Cercopithecus aethiops (Green monkey) (Givret).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecinae; Cercopithecus.
 CX NCBI_TaxID=9534;
 [1]

RP SEQUENCE FROM N.A.
 RA Mueller M.;
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Goernemann J.;
 RT "Cellular Interaction Partners of the HPV Minor Capsid Protein L2.";
 RL Thesis (2002), Department of Fakultät fuer Biologie,
 RL Universitaet Heidelberg, Heidelberg, Germany.
 DR EMBL: AJ437508; CAD26811.1; -.
 KW Nuclear protein.
 SQ SEQUENCE 222 AA; 25378 MW; 1DEDF23D388DAE04 CRC64;

Query Match 1.9%; Score 8; DB 6; Length 222;
 Best Local Similarity 100.0%; Pred. No. 63;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 RAALAA 17
 |||||
 Db 202 RAALAA 209

RESULT 26

O9Z5W3

ID Q925W3 PRELIMINARY; PRT; 238 AA.
 AC Q925W3;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Putative transcriptional regulatory protein Ohbr.
 GN OHBR.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OK NCBI_TaxId=287;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=142;
 RX MEDLINE=99240444; PubMed=10224014;
 RA Tsoi T.V., Plotnikova E.G., Cole J.R., Guerin W.F., Bagdasarian M.,
 RA Tiedje J.M.,
 RT "Cloning, expression, and nucleotide sequence of the Pseudomonas
 RT aeruginosa 142 ohb genes coding for oxygenolytic ortho dehalogenation
 RT of halobenzoates.";
 RL Appl. Environ. Microbiol. 65:2151-2162 (1999).
 CC -1- SIMILARITY: BELONGS TO THE ICRL FAMILY OF TRANSCRIPTIONAL
 CC REGULATORS.
 DR EMBL; AF121970; AAD20004.1; -;
 DR InterPro; IPR005471; HTH_ICLR.
 DR Pfam; PF01614; ICLR; 1.
 DR DNA-binding; Transcription; Transcription regulation.
 SQ SEQUENCE 238 AA; 25692 MW; 59CG78660D28A67E CRC64;

Query Match 1.9%; Score 8; DB 2; Length 238;
 Best Local Similarity 100.0%; Pred. No. 67;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 232 KLRQIAE 239
 |||||
 DB 152 KLRQIAE 159

RESULT 27
 ID Q96B43 PRELIMINARY; PRT; 253 AA.
 AC Q96B43;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Similar to methionyl aminopeptidase 2 (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OK NCBI_TaxId=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis, and Embryonic carcinoma;
 RA Straussberg R.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC016008; AAH16008.1; -;
 DR InterPro; IPR002468; MAP_2.
 DR InterPro; IPR001714; Methamino_PTase.
 DR InterPro; IPR000994; Peptidase_M24.
 DR Pfam; PF00557; Peptidase_M24; 1.
 DR PRINTS; PRO0559; MAPPTIDASE.
 DR TRGFAMS; TIGR00501; met_pdae_II; 1.
 DR PROSITE; PS01202; MAP_2; 1.
 DR AminoPeptidase.
 KW NON TER
 FT 1
 SQ SEQUENCE 253 AA; 28230 MW; B860E2D127731D6 CRC64;

Query Match 1.9%; Score 8; DB 4; Length 253;
 Best Local Similarity 100.0%; Pred. No. 70;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 203 NAGDTTVL 210
 |||||

DB 10 NAGDTTVL 17
 RESULT 28
 ID Q96295 PRELIMINARY; PRT; 290 AA.
 AC Q96295;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Putative secreted lipase.
 GN SC01735 OR SC111.24C.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinomycetales;
 OC Streptomycetaceae; Streptomycetaceae; Streptomyces.
 OK NCBI_TaxId=1902;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Saunders D., Harris D.;
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 RN (3)
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RX MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kleiser H.M., Denaplatte D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96 (1996).
 RN (4)
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RX MEDLINE=21996410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleiser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kleiser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabbinowitch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 RT coelicolor A3(2).";
 RL Nature 417:141-147 (2002).
 DR EMBL; AL939110; CAB50950.1; -;
 DR HSSP; Q05489; ITRH.
 DR InterPro; IPR002918; Lipase_2.
 DR InterPro; IPR000379; Ser_estr_ site.
 DR Pfam; PF01674; Lipase_2; 1.
 KW Complete proteome.
 SQ SEQUENCE 290 AA; 30496 MW; D56E8508350F7432 CRC64;

Query Match 1.9%; Score 8; DB 16; Length 290;
 Best Local Similarity 100.0%; Pred. No. 79;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 14 LAAVALV 21
 |||||
 DB 16 LAAVALV 23

RESULT 29
 ID Q926G6 PRELIMINARY; PRT; 296 AA.
 AC Q926G6;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Putative sugar uptake ABC transporter periplasmic solute-binding
DE protein.
GN RB0955 OR SMB2137.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Plantae; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396508; PubMed=11481431;
RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
RA Vorholter F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,
RA Golding B., Pehler A.,
RT "The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-
RT fixing endosymbiont Sinorhizobium meliloti."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
DR EMBL; AL603645; CAC49355.1; -
DR InterPro: IPR001761; PeriplasmBP/Lact.
DR Pfam; PF00532; Peripla_BP_like.1
KM Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 296 AA; 30625 MW; 43CB86F44829CF3 CRC64;

Query Match 1.9%; Score 8; DB 16; Length 296;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 AALAAVA 19
Db 216 AALAAVA 223

RESULT 30
Q9CH13 PRELIMINARY; PRT; 306 AA.
AC Q9CH13;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Transcription regulator.
GN YJFE OR L10928.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wincker P., Manger S., Jallion O., Malarne K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.,
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403."
RL Genome Res. 11:731-753(2001).
DR EMBL; AE006327; AAK05026.1; -
DR InterPro: IPR004474; LytR_cpSA Psr.
DR Pfam; PF03816; LytR_cpSA_Psr; 1.
KM Complete proteome.
SQ SEQUENCE 306 AA; 33657 MW; 2445A5CBA554CEBC CRC64;

Query Match 1.9%; Score 8; DB 16; Length 306;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 200 TLNAGDT 207
Db 202 TLNAGDT 209

RESULT 31
Q9SNL7

ID Q9SNL7 PRELIMINARY; PRT; 326 AA.
AC Q9SNL7;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Similar to magnesium-protoporphyrin IX methyltransferase (Putative
DE magnesium-protoporphyrin IX methyltransferase).
GN P0679C08.13.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nippobare (GA3) genomic DNA, chromosome 6, PAC
RT clone:P0493C11."
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nippobare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nippobare (GA3) genomic DNA, chromosome 6, PAC
RT clone:P0679C08."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP002542; BAB19374.1; -
DR Gramene; Q9SNL7; -
DR InterPro: IPR000551; SAM bind.
KM Methyltransferase; Transferase.
SQ SEQUENCE 326 AA; 34926 MW; F8C91146F1772278 CRC64;

Query Match 1.9%; Score 8; DB 10; Length 326;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 AALAAAV 18
Db 65 AALAAAV 72

RESULT 32
O67350 PRELIMINARY; PRT; 348 AA.
AC O67350;
DT 01-AUG-1998 (TReMBLrel. 07, Created)
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Cation efflux system (CZCS-1like).
GN CZCS2 OR AQ.1331.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VPS;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aulay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus."
RL Nature 392:353-358(1998).
DR EMBL; AE000735; AAC07317.1; -
DR InterPro: IPR005695; Membrane_fus2.
DR InterPro: IPR001950; TIF_SUI1.
DR TIGRFAMs; TIGR00999; 8a0102; 1.
DR PROSITE; PS01118; SUI1_1; 1.
KM Complete proteome.
SQ SEQUENCE 348 AA; 40570 MW; F4544EA6CA0635D CRC64;

Query Match 1.9%; Score 8; DB 16; Length 348;
 Best Local Similarity 100.0%; Pred. No. 93;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 358 LKDGKRV 365
 |||||
 DB 325 LKDGKRV 332

RESULT 33

Q8GRS3 PRELIMINARY; PRT; 390 AA.

AC Q8GRS3;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE OJ1477_F01.8 protein (P0047B07.31 protein).
 GN OJ1477_F01.8 OR P0047B07.31.
 OS Oryza sativa (japonica cultivar-group).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC SpERMatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 CC Ehrhacroidae; Oryzaceae; Oryza.
 OK NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare (GA) genomic DNA, chromosome 7, BAC clone:OJ1477_F01.8";
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsumoto T., Karayose Y.;
 RT "Oryza sativa nipponbare (GA) genomic DNA, chromosome 7, PAC clone:PO047B07.31";
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP003833; BAC15474.1; -
 DR EMBL; AP005184; BAC16140.1; -
 SQ SEQUENCE 390 AA; 44228 MW; BFD0C902BB861FAF CRC64;

Query Match 1.9%; Score 8; DB 10; Length 390;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ALAAVAL 20
 |||||
 DB 144 ALAAVAL 151

RESULT 34

Q9F241 PRELIMINARY; PRT; 394 AA.

AC Q9F241;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Putative membrane fusion protein.
 GN SMD.
 OS Xanthomonas maltophilia (Pseudomonas maltophilia) (Stenotrophomonas maltophilia).
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 CC Xanthomonadaceae; Stenotrophomonas.
 OK NCBI_TaxID=40324;
 RN [1].
 RP SEQUENCE FROM N.A.
 RC STRAIN=D457R;
 RX MEDLINE=20493115; PubMed=11036026;
 RA Alonso A., Martinez J.L.;
 RT "Cloning and characterization of SmeDF, a novel multidrug efflux pump from Stenotrophomonas maltophilia.";
 RL Antimicrob. Agents Chemother. 44:3079-3086(2000).
 DR EMBL; AJ252200; CAC14594.1; -

DR InterPro; IPR006143; HLYD.
 DR Pfam; PF00529; HLYD; 1.
 SQ SEQUENCE 394 AA; 40912 MW; D4234A037E47B53 CRC64;

Query Match 1.9%; Score 8; DB 2; Length 394;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 RITAPISG 189
 |||||
 DB 171 RITAPISG 178

RESULT 35

Q8P7C8 PRELIMINARY; PRT; 399 AA.

AC Q8P7C8;
 DT 01-OCT-2002 (TREMBlrel. 22, Created)
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Multidrug resistance protein.
 GN MEXA OR XCC2683.
 OS Xanthomonas campestris (pv. campestris).
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 CC Xanthomonadaceae; Xanthomonas.
 OK NCBI_TaxID=340;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 33913 / NCPPB 528;
 RX MEDLINE=22022145; PubMed=12024217;
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
 RA Quaggio R.B., Monteiro-Vitorcello C.B., Van Sluys M.A., Almeida N.F.,
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
 RA Camarotte G., Cannavan F., Cardozo J., Chamberg F., Clapina L.P.,
 RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
 RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
 RA Katayama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
 RA Locali E.C., Machado M.A., Madela A.M.B.N., Martinez-Rossi N.M.,
 RA Martins E.C., Mejdans J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
 RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
 RA Sphioia L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
 RA Trindade dos Santos M., Truffi D., Teal S.M., White F.F.,
 RA Serubal J.C., Kitajima J.P.;
 RT "Comparison of the genomes of two Xanthomonas pathogens with differing host specificities.";
 RL Nature 417:459-463(2002).
 DR EMBL; AE012380; AAM41955.1; -
 DR InterPro; IPR006143; HLYD.
 DR Pfam; PF00529; HLYD; 1.
 KW Complete proteome.
 SQ SEQUENCE 399 AA; 41511 MW; 59C1197DE6055D2D CRC64;

Query Match 1.9%; Score 8; DB 16; Length 399;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 122 AQATLAKA 129
 |||||
 DB 107 AQATLAKA 114

RESULT 36

Q8ZL01 PRELIMINARY; PRT; 408 AA.

AC Q8ZL01;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Putative cytoplasmic protein.
 GN YIDR OR STM3811.
 OS Salmonella typhimurium.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2 / SCS1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Milvanev E.,
 Ryan E., Sun H., Florea L., Miller W., Stenking T., Nhan M.,
 Waterston R., Wilson R.K.;
 RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium
 LT2.";
 RL Nature 413:852-856(2001).
 DR EMBL; AE008877; AL22670.1; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 408 AA; 45381 MW; 40454D052627FB08 CRC64;

Query Match 1.9%; Score 8; DB 16; Length 408;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 47 VGVTVHP 54
 |||||
 Db 64 VGVTVHP 71

RESULT 37
 Q822M0 PRELIMINARY; PRT; 408 AA.
 ID Q822M0;
 AC Q822M0;
 DT 01-MAR-2002 (TRENBLrel. 20, Created)
 DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
 DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
 DE Putative ATP/GTP-binding protein.
 GN STY3967.
 OS *Salmonella typhi*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
 Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
 Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 Feltham T., Hamlin N., Haque R., Hien T.T., Holroyd S., Jagerle K.,
 Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
 Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 Whitehead S., Barrell B.G.;
 RT "Complete genome sequence of a multiple drug resistant *Salmonella*
 enterica serovar Typhi CT18.";
 RL Nature 413:848-852(2001).
 DR EMBL; AL627280; CAD03183.1; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 408 AA; 45272 MW; 67EB07E7B9FF7B43 CRC64;

Query Match 1.9%; Score 8; DB 16; Length 408;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 47 VGVTVHP 54
 |||||
 Db 64 VGVTVHP 71

RESULT 38
 Q8XBZ9 PRELIMINARY; PRT; 416 AA.
 ID Q8XBZ9;
 AC Q8XBZ9;

DT 01-MAR-2002 (TRENBLrel. 20, Created)
 DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
 DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
 DE Orf, hypothetical protein.
 GN YIDR OR 25185 OR ECS4629.
 OS *Escherichia coli* O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 Postaf J., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 Grothbeck E.J., Davis N.W., Lim A., Dimantanta E.T., Potamousis K.,
 Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.";
 RL Nature 409:529-533(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / R1MD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Toke T.,
 Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,
 Kunara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*
 O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 DR EMBL; AB005600; AAC58891.1; -
 DR EMBL; AP002566; BAB38052.1; -
 KW Complete proteome.
 SQ SEQUENCE 416 AA; 46314 MW; 37D16EB952A88669 CRC64;

Query Match 1.9%; Score 8; DB 16; Length 416;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 47 VGVTVHP 54
 |||||
 Db 76 VGVTVHP 83

RESULT 39
 Q8FBW4 PRELIMINARY; PRT; 416 AA.
 ID Q8FBW4
 AC Q8FBW4;
 DT 01-MAR-2003 (TRENBLrel. 23, Created)
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE Hypothetical protein yidr.
 GN YIDR OR C4609.
 OS *Escherichia coli* O6.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=217992;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
 RX MEDLINE=22386234; PubMed=12471157;
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Rosach P.,
 Raeko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
 Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 of uropathogenic *Escherichia coli* O6.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
 DR EMBL; AB016769; AA083044.1; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 416 AA; 46325 MW; 529AC5CD6C08FC30 CRC64;

Query Match 1.9%; Score 8; DB 16; Length 416;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 VGVVTVHP 54
 |||||
 Db 76 VGVVTVHP 83

RESULT 40

Q8G7V0 PRELIMINARY; PRT; 442 AA.
 ID Q8G7V0
 AC Q8G7V0;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Possible soluble binding protein of ABC transporter.
 GN BL0141.
 OS Bifidobacterium longum.
 OC Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
 OC Bifidobacteriaceae; Bifidobacterium.
 OX NCBI_TaxID=216816;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=NC 2705;
 RX MEDLINE=22294977; PubMed=12381787;
 RA Schell M.A., Karimantzou M., Snel B., Vlianova D., Berger B.,
 RA Pessi G., Zwaalen M.-C., Desiere F., Bork P., Delley M.,
 RA Priamore R.D., Arigoni F.;
 RT "The genome sequence of Bifidobacterium longum reflects its adaptation
 RT to the human gastrointestinal tract."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
 DR EMBL: AE014630; AAN24000.1; -
 KW Complete proteome.
 SQ SEQUENCE 442 AA; 48635 MW; 1EBC9E9DACB45A9F CRC64;

Query Match 1.9%; Score 8; DB 16; Length 442;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 358 LKDGDKVV 365
 |||||
 Db 128 LKDGDKVV 135

Search completed: September 8, 2003, 14:10:47
 Job time : 70 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2003, 14:03:22 ; Search time 17 Seconds

(without alignments)
1139.706 Million cell updates/sec

Title: US-09-889-756A-2

Perfect score: 412
Sequence: 1 MAFYAFKAMRAAALAAVAL.....APOSQVOTASEAKTASEAE 412

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	89	21.6	412	1 MTRC_NEIGO	P43305 neisseria g
2	8	1.9	111	1 RL24_CHLMU	Q9pjms chlamydia m
3	8	1.9	111	1 RL24_CHLTR	P28337 chlamydia t
4	8	1.9	122	1 YM72_MYCTU	Q50691 mycobacteri
5	8	1.9	171	1 SFME_ECOLI	P38052 escherichia
6	8	1.9	284	1 TRPA_AZOB	Q91ag8 azospirillum
7	8	1.9	318	1 ARGB_BIFLO	P59295 bifidobacte
8	8	1.9	416	1 CHRA_PSEAE	P14285 pseudomonas
9	8	1.9	416	1 YIDR_ECOLI	P31455 escherichia
10	8	1.9	434	1 UL43_HSV11	P10227 herpes simp
11	8	1.9	465	1 GAA6_CHICK	Q90845 gallus gall
12	8	1.9	478	1 AMP2_HUMAN	P50579 homo sapien
13	8	1.9	478	1 AMP2_MOUSE	O08663 mus musculu
14	8	1.9	478	1 AMP2_RAT	P38062 rattus norv
15	8	1.9	890	1 ADHE_ECOLI	P17547 escherichia
16	8	1.7	91	1 MERP_SHIFL	P04129 shigella fl
17	8	1.7	98	1 YACB_RHISN	P55384 rhizobium s
18	8	1.7	106	1 VHSB_BPT3	P20322 bacterioph
19	8	1.7	106	1 YNUT_THOSH	Q01195 rhododact
20	8	1.7	111	1 RL24_CHLPM	Q927r8 chlamydia p
21	8	1.7	123	1 RL20_CHLMU	Q9184 chlamydia p
22	8	1.7	126	1 SYGB_NEIGO	Q50945 neisseria g
23	8	1.7	134	1 TYRT_STRGA	P55047 streptomyce
24	8	1.7	143	1 RS12_HORVU	Q9xh80 hordeum vul
25	8	1.7	148	1 SYRB_RHISN	P55361 rhizobium s
26	8	1.7	151	1 SYB2_RHIME	Q923q1 rhizobium m
27	8	1.7	151	1 SYB3_RHIME	P58346 rhizobium m
28	8	1.7	154	1 PSBU_CYACA	Q92gms cyanidium c
29	8	1.7	169	1 CYP2_ARATH	P35627 arabidopsis
30	8	1.7	169	1 YSR0_STRCO	Q92bg3 streptomyc
31	8	1.7	172	1 CYPH_CATRO	Q93613 catarranthu
32	8	1.7	177	1 IF3_ANASP	Q8yne3 anabaena sp
33	8	1.7	180	1 IF3_THIETN	Q8f9c2 thermoanaer

34	7	1.7	182	1 KRUC_SHEEP	P26372 ovis aries
35	7	1.7	183	1 RBFA_MYCTU	P71614 mycobacteri
36	7	1.7	185	1 IF3_RICPR	Q92d19 rickettsia
37	7	1.7	186	1 IF3_RICCN	Q92hk6 rickettsia
38	7	1.7	187	1 Y418_VIBCH	Q9kuv7 vibrio chol
39	7	1.7	197	1 PAT_ALCPA	P31668 alcaligenes
40	7	1.7	200	1 COAE_BRUME	O8ye21 brucella me
41	7	1.7	205	1 GSHC_BOVIN	P00435 bos taurus
42	7	1.7	218	1 ESM2_DROME	O97177 drosophila
43	7	1.7	226	1 OLE1_SOYBN	P29530 glycine max
44	7	1.7	232	1 SCOA_HELRY	Q92ies helicobacte
45	7	1.7	251	1 DABP_CAMUE	Q9pit2 campylobact
46	7	1.7	252	1 GUNX_PRUP	P38534 prunus pers
47	7	1.7	267	1 IF2A_METRA	Q8ty86 methanopyru
48	7	1.7	270	1 HMPH_HUMAN	Q03014 homo sapien
49	7	1.7	271	1 HMPH_MOUSE	P43120 mus musculu
50	7	1.7	278	1 PROI_BACSU	P54552 bacillus su
51	7	1.7	281	1 KPRS_HALNI	O9h488 halobacteri
52	7	1.7	283	1 AROE_METH	O26144 methanobact
53	7	1.7	286	1 YL33_ARCFU	Q26147 archaeoglob
54	7	1.7	302	1 ARGB_STRCL	Q91cs6 streptomyce
55	7	1.7	302	1 WR40_ARATH	O9esh7 arabidopsis
56	7	1.7	306	1 ARGB_STRCO	Q91ia3 streptomyc
57	7	1.7	312	1 HO2_RABIT	P43242 oryctolagus
58	7	1.7	322	1 GDC_RAT	P16261 rattus norv
59	7	1.7	332	1 DNJH_SYNY3	P50027 synecocyst
60	7	1.7	332	1 GDC_HUMAN	P16260 homo sapien
61	7	1.7	334	1 COBT_RHOCA	O52679 rhodobacter
62	7	1.7	334	1 YL28_SCHRO	Q10343 schizosacch
63	7	1.7	344	1 MRBW_MYCTU	O06212 mycobacteri
64	7	1.7	344	1 MURB_BUCAP	Q8Kae3 buchnera ap
65	7	1.7	366	1 VU1_HSV6U	O01349 human herpe
66	7	1.7	372	1 CYB_ELABA	O92250 elapha bair
67	7	1.7	372	1 FLGI_XANAC	O8p127 xanthomonas
68	7	1.7	372	1 FLGI_XANCP	O8p9c0 xanthomonas
69	7	1.7	372	1 LG6_PHACH	P50622 phanerocha
70	7	1.7	373	1 DDL_MYCTU	P50622 phanerocha
71	7	1.7	375	1 MURG_RHIL0	O988b3 rhizobium l
72	7	1.7	376	1 TGT_RHIME	O92dy4 rhizobium m
73	7	1.7	380	1 MURG_BRUME	O8y166 brucella me
74	7	1.7	386	1 CRYT_ERWHE	O01331 erythrina her
75	7	1.7	400	1 TCR8_PASMU	P01354 pasteurella
76	7	1.7	405	1 TCR5_ECOLI	Q07282 escherichia
77	7	1.7	406	1 METK_AERPE	O9y1b2 aeropyrum p
78	7	1.7	417	1 GLAI_PSEAE	O9htes pseudomonas
79	7	1.7	417	1 YAG1_YEAST	P39713 saccharomyc
80	7	1.7	418	1 GLA2_PSEAE	O91138 pseudomonas
81	7	1.7	437	1 YHAM_ECOLI	P42626 escherichia
82	7	1.7	441	1 VATH_ARATH	O91x65 arabidopsis
83	7	1.7	456	1 KIM8_HUMAN	O76015 homo sapien
84	7	1.7	458	1 A2AC_MOUSE	O01337 mus musculu
85	7	1.7	458	1 A2AC_RAT	P22086 rattus norv
86	7	1.7	469	1 ARLY_BURUL	P56161 burkholderi
87	7	1.7	470	1 ESCA_DROME	P25932 dirosophila
88	7	1.7	471	1 ARLY_FALSO	O8xwv7 raietonia s
89	7	1.7	471	1 KIM7_HUMAN	O76014 homo sapien
90	7	1.7	477	1 BAIG_EUBSE	P33369 eubacterium
91	7	1.7	494	1 GUN1_PERRA	P05522 persea amer
92	7	1.7	513	1 MATK_MAIZE	P48190 zea mays (m
93	7	1.7	515	1 ADCA_STRP3	O08484 streptococc
94	7	1.7	515	1 ADCA_STRP8	O8bln2 streptococc
95	7	1.7	519	1 ADCA_STRPY	O9a019 streptococc
96	7	1.7	519	1 DHAM_RAT	P11884 rattus norv
97	7	1.7	520	1 YW24_MYCTU	Q10509 mycobacteri
98	7	1.7	528	1 WR42_ARATH	O92ec3 arabidopsis
99	7	1.7	538	1 WR31_ARATH	O92ec3 arabidopsis
100	7	1.7	539	1 CH61_MYCTU	O59573 mycobacteri
101	7	1.7	542	1 P69_MYCPN	P73369 mycoplasma
102	7	1.7	556	1 SR54_CANAL	O42816 candida alb
103	7	1.7	568	1 GGT1_PIG	P20735 sus scrofa
104	7	1.7	569	1 AAKH_HUMAN	O9ugj0 homo sapien
105	7	1.7	590	1 HMAA_DROME	P25555 drosophila
106	7	1.7	597	1 CHIX_STR01	O05638 streptomyc


```

983 6 1.5 525 1 VIAA_HUMAN Q9H598 homo sapien
984 6 1.5 525 1 VIAA_MACFA Q95K62 macaca fasc
985 6 1.5 525 1 VIAA_MOUSE Q35633 mus musculu
986 6 1.5 525 1 VIAA_RAT Q35458 rattus norv
987 6 1.5 526 1 EXTL_RHIME Q92877 rhizobium m
988 6 1.5 526 1 MURE_COREF Q8RT55 corynebacte
989 6 1.5 526 1 NADB_MYCLE Q49617 mycobacteri
990 6 1.5 527 1 GSH1_PSEAE Q9HY66 pseudomonas
991 6 1.5 527 1 NADB_MYCTU Q06595 mycobacteri
992 6 1.5 527 1 PURI_MYCTU Q06626 mycobacteri
993 6 1.5 528 1 TCPO_MYCTU P39078 saccharomy
994 6 1.5 528 1 YC79_MYCTU Q11038 mycobacteri
995 6 1.5 529 1 SP15_STRGR P19471 streptomyce
996 6 1.5 529 1 VGLF_SVS P04849 simian viru
997 6 1.5 530 1 UDB2_RAT P08541 rattus norv
998 6 1.5 531 1 YP60_METTM P14933 methanobact
999 6 1.5 532 1 YMO0_YEAST Q04458 saccharomy
1000 6 1.5 533 1 APO2_STRCO Q04943 streptomyce

```

ALIGNMENTS

```

RESULT 1
MTRC_NEIGO STANDARD; PRT; 412 AA.
ID MTRC_NEIGO
AC P43505;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Membrane fusion protein mtrc precursor.
GN MTRC.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxId=485;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=FA19;
RX MEDLINE=94254732; PubMed=8196548;
RA Pan W., Spratt B.G.;
RT "Regulation of the permeability of the gonococcal cell envelope by
the mtr system.";
RL Mol. Microbiol. 11:769-775(1994).
CC -1- FUNCTION: CELL MEMBRANE LIPOPROTEIN, INVOLVED IN CELL MEMBRANE
PERMEABILITY TO HYDROPHOBIC COMPOUNDS SUCH AS ANTIBIOTICS, DYES
AND DETERGENTS.
CC -1- SUBCELLULAR LOCATION: Attached to the inner membrane by a lipid
anchor (Potential).
CC -1- SIMILARITY: BELONGS TO THE ACRA/ACRE FAMILY.
-----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation-
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC CC EMBL: U14993; AAA80193.1; -
CC DR InterPro; IPR006143; H1yD.
CC DR Pfam; PF00529; H1yD; 1.
CC DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
CC KW Signal; Inner membrane; Lipoprotein.
CC FT CHAIN 1 24 POTENTIAL.
CC FT LIPID 25 412 MEMBRANE FUSION PROTEIN MTRC.
CC FT LIPID 25 25 N-ACYL DIGLYCERIDE (POTENTIAL).
CC SQ SEQUENCE 412 AA; 42773 MW; 979FAFBCFAE321BA CRC64;

```

Query Match 21.6%; Score 89; DB 1; Length 412;
 Best Local Similarity 100.0%; Pred. No. 1e-77;
 Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 22 LSSCGKGDADAAQGGPAPREAPAPVGVVTHPQTVALTELPGRLBSLRADVAQVGG 81
Db 22 LSSCGKGDADAAQGGPAPREAPAPVGVVTHPQTVALTELPGRLBSLRADVAQVGG 81
Qy 82 IIOKRLFOEGSVYRAGPFLYQIDSTVEA 110
Db 82 IIOKRLFOEGSVYRAGPFLYQIDSTVEA 110

```

```

RESULT 2
RL24_CHLMTU STANDARD; PRT; 111 AA.
ID RL24_CHLMTU
AC Q9PUM5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 50S ribosomal protein L24.
GN RPLX OR TC0804.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxId=83560;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=Mopn / N199;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Ullerbach T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -1- FUNCTION: THIS PROTEIN IS FOUND IN THE RIBONUCLEOPROTEIN CORE AND
IS INVOLVED IN THE EARLY ASSEMBLY OF THE 50S SUBUNIT. IT IS NOT
INVOLVED IN THE FUNCTIONS OF THE MATURE 50S SUBUNIT (BY
SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE L24P FAMILY OF RIBOSOMAL PROTEINS.
-----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation-
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC CC EMBL: AE002347; AAP39607.1; -
CC DR PIR; D81664; D81664.
CC DR PIR; E42645; E42645.
CC DR TIGR; TC0804; -
CC DR InterPro; IPR005824; KOW.
CC DR InterPro; IPR00646; KOW_sub.
CC DR InterPro; IPR003256; Ribosomal_L24.
CC DR InterPro; IPR005825; Ribosomal_L24_26.
CC DR Pfam; PF00467; KOW; 1.
CC DR ProDom; PD001677; Ribosomal_L24; 1.
CC DR SMART; SMO0739; KOW; 1.
CC DR TIGRFAMs; TIGR01079; rplX_bact; 1.
CC DR PROSITE; PS01108; RIBOSOMAL_L24; 1.
CC KW Ribosomal protein; Complete proteome.
CC SQ SEQUENCE 111 AA; 12608 MW; 2F02863BA68C13F5 CRC64;

```

Query Match 1.9%; Score 8; DB 1; Length 111;
 Best Local Similarity 100.0%; Pred. No. 3.9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 362 DKVVEGI 369
 |||||
 Db 32 DKVVEGI 39

RESULT 3

```

RL24 CHLTR
ID RL24_CHLTR STANDARD; PRT; 111 AA.
AC P28537; 084523;
DT 01-DEC-1992 (Rel. 24, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 50S ribosomal protein L24.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=L2/434/Bu;
RX MEDLINE=92138612; PubMed=1735714;
RA Kaul R., Gray G.J., Koehnke N.R., Gu L.J.;
RT "Cloning and sequence analysis of the Chlamydia trachomatis spc
RT ribosomal protein gene cluster."
RL J. Bacteriol. 174:1205-1212(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UW-3/Cx;
RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kaiman S., Lammell C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatunov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis."
RL Science 282:754-759(1998).
CC -1- FUNCTION: THIS PROTEIN IS FOUND IN THE RIBONUCLEOPROTEIN CORE AND
CC IS INVOLVED IN THE EARLY ASSEMBLY OF THE 50S SUBUNIT. IT IS NOT
CC INVOLVED IN THE FUNCTIONS OF THE MATURE 50S SUBUNIT. (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE L24P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M80325; AAA23173.1; -
DR EMBL; AE001323; AAC68118.1; -
DR InterPro; IPR005824; KOM.
DR InterPro; IPR006646; KOM_sub.
DR InterPro; IPR003256; Ribosomal_L24.
DR InterPro; IPR005825; Ribosomal_L24_26.
DR Pfam; PF00467; KOM; 1.
DR ProDom; PD001677; Ribosomal_L24; 1.
DR SMART; SM00739; KOM; 1.
DR TIGFAMS; TIGR01079; tpxX_bact; 1.
DR PROSITE; PS01108; RIBOSOMAL_L24; 1.
KW Ribosomal protein; Complete proteome.
FT VARIANT 85 R -> T (IN SEROVAR L2).
FT VARIANT 105 S -> L (IN SEROVAR L2).
SQ SEQUENCE 111 AA; 12637 MW; 2F02805BC08C0165 CRC64;

Query Match 1.9%; Score 8; DB 1; Length 111;
Best Local Similarity 100.0%; Pred. No. 3.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 362 DKVVEGI 369
Db 32 DKVVEGI 39

RESULT 4
ID YM72_MYCTU STANDARD; PRT; 122 AA.
AC Q50691;
DT 01-NOV-1997 (Rel. 35, Created)

```

```

DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein RV2272.
GN RV2272 OR MT2333 OR MTCY339.38C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=96295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolony J.F., Nelson W.C., Umayam L.A., Ernolava M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Kouri H., Gill J., Mikus A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC
CC -1- SIMILARITY: TO E. COLI YIDH.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z77163; CAB00963.1; -
DR EMBL; AE007076; AAK4616.1; -
DR PIR; D70730; D70730.
DR TIGR; MT2333; -
DR Tuberculist; RV2272; -
DR InterPro; IPR003807; DUF202.
DR Pfam; PF02656; DUF202; 1.
DR Hypothetical protein; Transmembrane; Complete proteome.
KW TRANSMEM 33 POTENTIAL.
FT TRANSMEM 58 78 POTENTIAL.
FT TRANSMEM 97 117 POTENTIAL.
SQ SEQUENCE 122 AA; 12994 MW; 23A53754264687A7 CRC64;

Query Match 1.9%; Score 8; DB 1; Length 122;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 14 LAAAVLV 21
Db 37 LAAAVLV 44

RESULT 5
ID SPMF_ECOLI STANDARD; PRT; 171 AA.
AC P38052; P75716; P77079;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)

```

DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Fimbrial-like protein sfmf precursor.
 GN SFMF OR B0534.
 OS Escherichia coli.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; Escherichia.
 CC NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RP Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,
 RA Federfpiel N., Hyman R., Kaiman S., Komp C., Kurdi O., Lew H.,
 RA Lin D., Natch A., Oefner P., Roberts D., Schramm S., Davis R.W.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBD databases.
 RN [3]
 RP SEQUENCE OF 119-171 FROM N.A.
 RP STRAIN=K12;
 RX MEDLINE=90220507; PubMed=2183007;
 RA Muramatsu S., Mizuno T.;
 RT "Nucleotide sequence of the region encompassing the int gene of a
 RT cryptic prophage and the dna Y gene flanked by a curved DNA sequence
 RT of Escherichia coli K12.";
 RL Mol. Gen. Genet. 220:325-328(1990).
 RN [4]
 RP IDENTIFICATION
 RP MEDLINE=95075659; PubMed=7984428;
 RA Borodovsky M., Rudd K.E., Koonin E.V.;
 RT "Intrinsic and extrinsic approaches for detecting genes in a
 RT bacterial genome.";
 RL Nucleic Acids Res. 22:4756-4767(1994).
 CC -1- SIMILARITY: BELONGS TO THE FIMB/PAPA FAMILY OF FIMBRIA PROTEINS.
 CC -1- SIMILARITY: STRONG, TO S.TYPHIMORIUM FIMF.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE000159; AAC73636.1; -;
 DR EMBL; U82598; AAB40732.1; ALT_INIT.
 DR EMBL; X51662; -; NOT_ANNOTATED_CDS.
 DR PIR; E64785; E64785.
 DR Ecogen; EG12388; sfmf.
 DR InterPro; IPR000259; Fimbrial.
 DR Pfam; PF00419; Fimbrial; 1.
 KM Fimbria; Signal; Complete proteome.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 171 FIMBRIAL-LIKE PROTEIN SFMF.
 FT DISULFID 40 80 POTENTIAL.
 SQ SEQUENCE 171 AA; 18222 MW; 59230981242AIDCA CRC64;

Query Match 1.9%; Score 8; DB 1; Length 171;
 Best Local Similarity 100.0%; Pred. No. 5.5;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 201 LINAAGDTT 208
 |||||
 DB 61 LINAAGDTT 68

RESULT 6

TRPA_AZOBR STANDARD; PRT; 284 AA.
 ID TRPA_AZOBR
 AC O9LAG9;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Tryptophan synthase alpha chain (EC 4.2.1.20).
 GN TRPA.
 OS Acetivillium brasiliense.
 CC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
 CC Rhodospirillaceae; Azospirillum.
 CC NCBI_TaxID=192;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP STRAIN=SP7;
 RX MEDLINE=20541289; PubMed=11092742;
 RA Dosselare F., Lambrecht M., Vanderleyden J.;
 RT "Isolation and sequence analysis of the trpA gene cluster, encoding
 RT tryptophan synthase, from Azospirillum brasiliense.";
 RL DNA Seq. 11:287-293(2000).
 CC -1- FUNCTION: The alpha subunit is responsible for the aldo cleavage
 CC of indoleglycerol phosphate to indole and glyceraldehyde 3-
 CC phosphate.
 CC -1- CATALYTIC ACTIVITY: L-serine + 1-(indol-3-yl)glycerol 3-phosphate
 CC = L-tryptophan + glyceraldehyde 3-phosphate.
 CC -1- PATHWAY: Tryptophan biosynthesis; fifth (last) step.
 CC -1- SUBUNIT: Tetramer of two alpha and two beta chains (By
 CC similarity).
 CC -1- SIMILARITY: BELONGS TO THE TRPA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF139661; AAF61458.1; -;
 DR HSRP; P00929; 2MSY.
 DR HAMAP; MF_00131; -; 1.
 DR InterPro; IPR003009; FMN_enzyme.
 DR InterPro; IPR002028; trp_synthaseA.
 DR Pfam; PF00290; trp_synthA; 1.
 DR ProDom; PD001535; Trp_synthaseA; 1.
 DR TIGRFAMs; TIGR00262; trpA; 1.
 DR PROSITE; PS00167; TRP_SYNTHASE_ALPHA; 1.
 KW Tryptophan biosynthesis; Lyase.
 SQ SEQUENCE 284 AA; 29351 MW; EBB12C8727BF1C57 CRC64;

Query Match 1.9%; Score 8; DB 1; Length 284;
 Best Local Similarity 100.0%; Pred. No. 8.3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 370 SIAGITGA 377
 |||||
 DB 188 SIAGITGA 195

RESULT 7
 ARGB_BIFLO STANDARD; PRT; 318 AA.
 ID ARGB_BIFLO
 AC P59255;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Acetylglutamate kinase (EC 2.7.2.8) (NAG kinase) (AGX) (N-acetyl-L-
 DE glutamate 5-phosphotransferase).
 GN ARGB OR BL1062.
 OS Bifidobacterium longum.
 CC Bacteria; Actinobacteria; Actinobacteriales; Bifidobacteriales;
 CC Bifidobacteriaceae; Bifidobacterium.
 CC NCBI_TaxID=216816;

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCC 2705;
RX MEDLINE=22294977; PubMed=12381787;
RA Schell M.A., Karmirantzou M., Snell B., Vilanova D., Berger B.,
RA Pessi G., Zehren M.-C., Desiere F., Bork P., Delley M.,
RA Pridmore R.D., Arigoni F.;
RT "The genome sequence of Bifidobacterium longum reflects its adaptation
RT to the human gastrointestinal tract."
RT Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
CC -1- CATALYTIC ACTIVITY: ATP + N-acetyl-L-glutamate = ADP + N-acetyl-L-
CC glutamate 5-phosphate.
CC -1- PATHWAY: Arginine biosynthesis; second step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- SIMILARITY: Belongs to the acetylglutamate kinase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE014729; AAN24870.1; -.
DR HAMAP; MF_00082; -.
DR Pfam; PF00696; aakkinase; 1.
DR TIGRFAMs; TIGR00761; argB; 1.
KW Arginine biosynthesis; Transferase; Kinase; Complete proteome.
FT ACT_SITE 45 45 BY SIMILARITY.
FT ACT_SITE 264 264 BY SIMILARITY.
FT SITE 100 100 SUBSTRATE BINDING (BY SIMILARITY).
FT SITE 102 102 SUBSTRATE BINDING (BY SIMILARITY).
FT SITE 203 203 SUBSTRATE BINDING (BY SIMILARITY).
SQ SEQUENCE 318 AA; 33818 MW; EC352CB1D32E1B7 CRC64;

Query Match 1.9%; Score 8; DB 1; Length 318;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 AALAAAV 18
DB 210 AALAAAV 217

RESULT 8
CHRA_PSEAE STANDARD; PRT; 416 AA.
AC P14285;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Chromate transport protein.
GN CHRA.
OS Pseudomonas aeruginosa.
OG Plasmid pJM505.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90094230; PubMed=2152903;
RA Ceravantes C., Ohake H., Chu L., Misra T.K., Silver S.;
RT "Cloning, nucleotide sequence, and expression of the chromate
RT resistance determinant of Pseudomonas aeruginosa plasmid pJM505."
RL J. Bacteriol. 172:287-291(1990).
CC -1- FUNCTION: THIS PROTEIN REDUCES CHROMATE ACCUMULATION AND IS
CC ESSENTIAL FOR CHROMATE RESISTANCE.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (potential).
CC -1- INDUCTION: By chromate.
CC -1- SIMILARITY: 29% IDENTITY TO ALCALIGENES EUTROPHUS CHROMATE
CC TRANSPORT PROTEIN CHRA.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M29034; AAA88432.1; -.
DR InterPro; IPR003370; Chromate_transp.
DR Pfam; PF02417; Chromate_transp; 2.
DR TIGRFAMs; TIGR00937; 2A51; 1.
KW Transport; Plasmid; Transmembrane; Chromate resistance.
FT TRANSMEM 73 93 POTENTIAL.
FT TRANSMEM 99 119 POTENTIAL.
FT TRANSMEM 128 148 POTENTIAL.
FT TRANSMEM 164 184 POTENTIAL.
FT TRANSMEM 204 224 POTENTIAL.
FT TRANSMEM 237 257 POTENTIAL.
FT TRANSMEM 283 303 POTENTIAL.
FT TRANSMEM 308 328 POTENTIAL.
FT TRANSMEM 341 361 POTENTIAL.
FT TRANSMEM 371 391 POTENTIAL.
FT TRANSMEM 395 415 POTENTIAL.
SQ SEQUENCE 416 AA; 44060 MW; 19F46F4D3B0D2479 CRC64;

Query Match 1.9%; Score 8; DB 1; Length 416;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ALAAVAL 20
DB 213 ALAAVAL 220

RESULT 9
YIDR_ECOLI STANDARD; PRT; 416 AA.
ID YIDR_ECOLI
AC P31455; P76734;
DT 01-JUL-1993 (Rel. 26, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein yidR.
GN YIDR OR B3689.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93315143; PubMed=7686882;
RA Burland V.D., Plunkett G. III, Daniels D.L., Blattner F.R.;
RT "DNA sequence and analysis of 136 kilobases of the Escherichia coli
RT genome: organizational symmetry around the origin of replication."
RL Genomics 16:551-561(1993).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L10328; AAA62041.1; -.
DR EMBL; AE000446; AAC76712.1; -.
DR PIR; B65171; B65171.
DR Ecocyc; Egl1713; yidR.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 416 AA; 46318 MW; 47E36360CC89A8B7 CRC64;

Query Match 1.9%; Score 8; DB 1; Length 416;

```

Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 VGVWTVHP 54
DB 76 VGVWTVHP 83

RESULT 10

UL43_HSV11 STANDARD; PRT; 434 AA.

AC P10227; 01-MAR-1989 (Rel. 10, Created)

DT 01-MAR-1989 (Rel. 10, Last sequence update)

DE 01-DEC-1992 (Rel. 24, Last annotation update)

Membrane protein UL43.

OS Herpes simplex virus (type 1 / strain 17).

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

CC Alphaherpesvirinae; Simplexvirus.

OX NCBI_TaxID=10299;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=88274327; PubMed=2839594;

RA McGeech D.J., Dalrymple M.A., Davison A.J., Dolan A., Frame M.C.,

McNab D., Perry L.J., Scott J.E., Taylor P.,

"The complete DNA sequence of the long unique region in the genome of

herpes simplex virus type 1."

RT J. Gen. Virol. 69:1531-1574(1988).

CC -1- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED OR TRANSMEMBRANE

CC PROTEIN (POTENTIAL).

CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL43,

EHV-1 17, AND VZV 15.

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

DR EMBL; X14112; CAA32306.1; -

DR PIR; G30088; WMBEP3.

DR Pfam; PF05072; Herpes_UL43; 1.

KW Membrane.

SQ SEQUENCE 434 AA; 44908 MW; 2ABCCSE02P96F0D2 CRC64;

Query Match 1.9%; Score 8; DB 1; Length 434;

Best Local Similarity 100.0%; Pred. No. 12;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 AAALAAV 18

DB 199 AAALAAV 206

RESULT 11

GAA6_CHICK STANDARD; PRT; 465 AA.

ID GAA6_CHICK 090845;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DE 28-FEB-2003 (Rel. 41, Last annotation update)

Gamma-aminobutyric-acid receptor alpha-6 subunit precursor (GABA(A)

receptor).

GN GABRA6.

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.

OX NCBI_TaxID=9031;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Embryonic brain;

RX MEDLINE=96373714; PubMed=8780005;

RA Bahn S., Harvey R.J., Dalrymple M.G., Wisden W.;

"Conservation of gamma-aminobutyric acid type A receptor alpha 6

subunit gene expression in cerebellar granule cells."

RT J. Neurochem. 66:1810-1818(1996).

CC -1- FUNCTION: GABA, THE MAJOR INHIBITORY NEUROTRANSMITTER IN THE

VERTEBRATE BRAIN, MEDIATES NEURONAL INHIBITION BY BINDING TO THE

GABA/BENZODIAZEPINE RECEPTOR AND OPENING AN INTEGRAL CHLORIDE

CHANNEL.

CC -1- SUBUNIT: GENERALLY PENTAMERIC. THERE ARE FIVE TYPES OF GABA(A)

CC RECEPTOR CHAINS: ALPHA, BETA, GAMMA, DELTA, AND RHO.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein.

CC -1- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNEL FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

DR EMBL; X94343; CAA64069.1; -

DR InterPro; IPR005436; GABRA6_receptor.

DR InterPro; IPR006029; Neu channel memb.

DR InterPro; IPR006202; Neu_chan_LBD.

DR InterPro; IPR006201; Neu_chan_LBD.

DR Pfam; PF02931; Neu_chan_LBD; 1.

DR Pfam; PF02932; Neu_chan_memb; 1.

DR PRINTS; PR01619; GABARALPHA6.

DR PRINTS; PR00252; NRIONCHANNEL.

DR TIGRFAMs; TIGR00860; LIC; 1.

DR PROSITE; PS00236; NEURON_ION_CHANNEL; 1.

KW Postsynaptic membrane; Ionic channel; Glycoprotein; Signal;

KW Transmembrane; Multigene family.

FT SIGNAL 1 19

FT CHAIN 20 465

FT POTENTIAL.

FT GAMMA-AMINOBUTYRIC-ACID RECEPTOR ALPHA-6

FT SUBUNIT.

FT EXTRACELLULAR (PROBABLE).

FT TRANSMEM 243 264

FT TRANSMEM 269 290

FT TRANSMEM 302 324

FT DOMAIN 325 421

FT TRANSMEM 422 443

FT CARBOHYD 31 31

FT CARBOHYD 128 128

FT CARBOHYD 141 141

FT DISULFID 156 170

FT BY SIMILARITY.

SQ SEQUENCE 465 AA; 51717 MW; 969061807FBCDD09 CRC64;

Query Match 1.9%; Score 8; DB 1; Length 465;

Best Local Similarity 100.0%; Pred. No. 12;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 RAALAAA 17

DB 341 RAALAAA 348

RESULT 12

AMP2_HUMAN STANDARD; PRT; 478 AA.

ID AMP2_HUMAN P50579;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DE 15-SEP-2003 (Rel. 42, Last annotation update)

DE Methionine aminopeptidase 2 (EC 3.4.11.18) (MetAP 2) (Peptidase M 2)

GN (Initiation factor 2 associated 67 kDa glycoprotein) (p67) (p67e1f2).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Carnivora; Hominoidea; Homo.
 OK NCBI_taxid=9606;
 [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=95372350; PubMed=7644482;
 RA Artin S.M., Kendall R.L., Hall L., Weaver L.H., Stewart A.E.,
 RA Matthews B.W., Bradshaw R.A.;
 RT "Eukaryotic methionyl aminopeptidases: two classes of cobalt-dependent
 enzymes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:7714-7718(1995).
 [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=95178556; PubMed=7873610;
 RA Li X., Chang Y.;
 RT "Molecular cloning of a human complementary DNA encoding an
 initiation factor 2-associated protein (p67).";
 RL Biochim. Biophys. Acta 1260:333-336(1995).
 [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lymph;
 RX MEDLINE=22386257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
 RA Brownstein M.J., Uscin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwen P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Small D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 [4]
 RP X-RAY CRYSTALLOGRAPHY (1.8 ÅNGSTROMS) OF 109-478.
 RX MEDLINE=99030697; PubMed=9812898;
 RA Liu S., Widom J., Kemp C.W., Crews C.M., Clardy J.;
 RT "Structure of human methionine aminopeptidase-2 complexed with
 fumagillin.";
 RL Science 282:1324-1327(1998).
 CC -!- FUNCTION: REMOVES THE AMINO-TERMINAL METHIONINE FROM NASCENT
 PROTEINS.
 CC -!- CATALYTIC ACTIVITY: Release of N-terminal amino acids,
 preferentially methionine, from peptides and arylamides.
 CC -!- COFACTOR: COBALT; BINDS 2 IONS PER SUBUNIT (By similarity).
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M24C.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U29607; AA82930.1; -;
 DR EMBL; U13261; AAC63402.1; -;
 DR EMBL; BC013782; AA13782.1; -;
 DR PIR; S52112; DPHUM2.
 DR PDB; 1B59; 14-JAN-00.
 DR PDB; 1B6A; 18-FEB-00.
 DR PDB; 1BN5; 31-JUL-99.
 DR PDB; 1BOA; 01-AUG-99.

DR MEROPS; M24.002; -;
 DR Genew; HGNC:16672; METAP2.
 DR MIM; 601870; -;
 DR GO; GO:0004239; F:methionyl aminopeptidase activity; TAS.
 DR GO; GO:0006464; P:protein modification; TAS.
 DR GO; GO:0006445; P:regulation of translation; TAS.
 DR InterPro; IPR002468; MAP_2.
 DR InterPro; IPR001714; Methamino_Prase.
 DR InterPro; IPR000994; Peptidase M24.
 DR Pfam; PF00557; Peptidase M24; 1.
 DR PRINTS; PR00599; MAPPEPTIDASE.
 DR TIGRfams; TIGR00501; met_pdae_II; 1.
 DR PROSITE; PS01202; MAP_2; 1.
 KW Hydroxylase; Aminopeptidase; Cobalt; 3D-structure.
 FT DOMAIN 36 46
 FT PROSITE; ARG/LYS-RICH (BASIC).
 FT ACT_SITE 98 93
 FT POLY-LYS.
 FT ACT_SITE 231 231
 FT METAL 251 251
 FT METAL 262 262
 FT METAL 331 331
 FT METAL 364 364
 FT METAL 459 459
 FT METAL 120 123
 FT METAL 125 126
 FT METAL 133 135
 FT METAL 154 161
 FT METAL 162 162
 FT METAL 163 186
 FT METAL 189 190
 FT METAL 192 192
 FT METAL 193 208
 FT METAL 209 209
 FT METAL 211 211
 FT METAL 212 214
 FT METAL 215 225
 FT METAL 226 227
 FT METAL 228 230
 FT METAL 236 237
 FT METAL 241 241
 FT METAL 244 245
 FT METAL 248 256
 FT METAL 257 258
 FT METAL 259 267
 FT METAL 271 273
 FT METAL 274 290
 FT METAL 293 294
 FT METAL 296 296
 FT METAL 297 309
 FT METAL 310 310
 FT METAL 312 315
 FT METAL 316 317
 FT METAL 318 321
 FT METAL 323 324
 FT METAL 326 327
 FT METAL 329 334
 FT METAL 335 336
 FT METAL 337 337
 FT METAL 343 345
 FT METAL 354 354
 FT METAL 357 358
 FT METAL 360 368
 FT METAL 375 376
 FT METAL 382 385
 FT METAL 387 388
 FT METAL 397 409
 FT METAL 410 412
 FT METAL 415 416
 FT METAL 417 422
 FT METAL 423 424
 FT METAL 429 437
 FT METAL 438 439
 FT METAL 441 444


```

DR PROSITE, PS01202; MAP 2; 1.
KW Hydrolyase; Aminopeptidase; Cobalt; Glycoprotein.
FT DOMAIN 36 46 ARG/LYS-RICH (BASIC).
FT DOMAIN 82 93 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 98 106 POLY-LYS.
FT METAL 251 251 COBALT 2 (By similarity).
FT METAL 262 262 COBALT 1 AND 2 (By similarity).
FT METAL 331 331 COBALT 1 (By similarity).
FT METAL 364 364 COBALT 1 (By similarity).
FT METAL 459 459 COBALT 1 AND 2 (By similarity).
FT CONFLICT 464 478 LRPCKEVSRGDDY -> CAQPVKLSAEHTIKT (IN REF. 1).
SQ SEQUENCE 478 AA; 53052 MW; BEIC0E91EDCB3D74 CRC64;

Query Match 1.9%; Score 8; DB 1; Length 478;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 203 NAGDTTVL 210
Db 235 NAGDTTVL 242

RESULT 15
ADHE_ECOLI STANDARD; PRT; 890 AA.
AC ADHE_ECOLI 17547;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Aldehyde-alcohol dehydrogenase [includes: Alcohol dehydrogenase
DE (EC 1.1.1.1) (ADH); Acetaldehyde dehydrogenase [acetylating]
DE (EC 1.2.1.10) (ACDH); Pyruvate-formate-lyase deactivase (PFL
DE deactivase)].
GN ADHE OR ANA OR B1241 OR Z2016 OR ECS1741.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.
RA MEDLINE=90152365; PubMed=2695398;
RA Goodlove P.E., Cunningham P.R., Parker J., Clark D.P.;
RT "Cloning and sequence analysis of the fermentative alcohol-
RT dehydrogenase-encoding gene of Escherichia coli.";
RL Gene 85:209-214(1989).
RN [2]
RP SEQUENCE FROM N.A., SEQUENCE OF 1-6, AND CHARACTERIZATION.
RC STRAIN=K12;
RC MEDLINE=91200315; PubMed=2015910;
RA Kessler D., Leibrecht I., Knappe J.;
RT "Pyruvate-formate-lyase-deactivase and acetyl-CoA reductase
RT activities of Escherichia coli reside on a plasmidic protein particle
RT encoded by adhE.";
RL FEBS Lett. 281:59-63(1991).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RC MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RC MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,

```

```

RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA Samesi G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RC MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blatner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RC MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ono Sudo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasekawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RN [7]
RP SEQUENCE OF 848-890 FROM N.A.
RC STRAIN=K12;
RC MEDLINE=95291445; PubMed=7773397;
RA Danchin A., Krin E.;
RT "Filling the gap between hns and adhE in Escherichia coli K12.";
RL Microbiology 141:959-960(1995).
RN [8]
RP SEQUENCE OF 1-20.
RC MEDLINE=94359415; PubMed=7521508;
RA Yamato M., Takahashi Y., Tomotake H., Ota F., Hirota K., Yamaguchi K.;
RT "Monoclonal antibodies to spiroxin of Yersinia enterocolitica and
RT analysis of the localization of spiroxin by use of them.";
RL Microbiol. Immunol. 38:177-182(1994).
CC -1- FUNCTION: THIS ENZYME HAS THREE ACTIVITIES: ADH, ACDH, AND PFL-
CC DEACTIVASE.
CC -1- FUNCTION: PFL DEACTIVASE CATALYZES THE QUENCHING OF THE PYRUVATE-
CC FORMATE-LYASE CATALYST IN AN IRON, NAD, AND COA DEPENDENT
CC REACTION.
CC -1- CATALYTIC ACTIVITY: An alcohol + NAD(+) = an aldehyde or ketone +
CC NADH.
CC -1- CATALYTIC ACTIVITY: Acetaldehyde + CoA + NAD(+) = acetyl-CoA +
CC NADH.
CC -1- COPACITOR: IRON ACTIVATED.
CC -1- SUBUNIT: SEEMS TO FORM A ROD SHAPED POLYMER COMPOSED OF ABOUT 40
CC IDENTICAL SUBUNITS.
CC -1- INDUCTION: UNDER ANAEROBIC CONDITIONS IN THE ABSENCE OF NITRATE.
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE IRON-
CC CONTAINING ALCOHOL DEHYDROGENASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; X59263; CA44195.1; -
CC EMBL; M33504; AAA23420.1; -
CC EMBL; AE000222; AAC74323.1; -
CC EMBL; D90759; BAA36121.1; -
CC EMBL; D90852; BAA16034.1; -
CC EMBL; AE005341; AAG56096.1; -

```



```

OS Rhizobium sp. (strain NGR234).
OG Plasmid sym pNG234a.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
OX NCB1_TaxID=394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97305956; PubMed=9163424;
RA Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA Perret X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes."
RL Nature 387:394-401(1997).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: NONE OBVIOUS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; A8000067; AAB91632.1; -
DR PIR; T28629; T28629.
KW Hypothetical protein; Plasmid; Transmembrane.
FT TRANSMEM 2 POTENTIAL.
FT TRANSMEM 70 90 POTENTIAL.
SQ SEQUENCE 98 AA; 10241 MW; B3733265061385 CRC64;

Query Match
Best Local Similarity 1.7%; Score 7; DB 1; Length 98;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 AAAAIAA 17
Db 73 AAAAIAA 79

RESULT 18
VHSB_BPT3
ID VHSB_BPT3 STANDARD; PRT; 106 AA.
AC P20322;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE Host specificity protein B.
GN 7.3.
OS Bacteriophage T3.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
OC T7-like viruses.
OX NCB1_TaxID=10759;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Luria;
RC MEDLINE=90133923; PubMed=2614843;
RA Beck P.J., Gonzalez S., Ward C.L., Molineux I.J.;
RT "Sequence of bacteriophage T3 DNA from gene 2.5 through gene 9."
RL J. Mol. Biol. 210:687-701(1989).
CC -1- FUNCTION: AFFECTS THE HOST RANGE (PROBABLE).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X17255; CAA35151.1; -
DR PIR; S07520; S07520.
RX MEDLINE=99206606; PubMed=10192388;
SQ SEQUENCE 106 AA; 10842 MW; 5006890DF38A73A3 CRC64;

```

```

Query Match
Best Local Similarity 1.7%; Score 7; DB 1; Length 106;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 EAPAPV 47
Db 48 EAPAPV 54

RESULT 19
YNU_RHOSH
ID YNU_RHOSH STANDARD; PRT; 106 AA.
AC Q01195;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Hypothetical 10.8 kDa protein in nifu 5' region (ORF 1).
OS Rhodobacter sphaeroides (Rhodospirillum rubrum sphaeroides).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
OC Rhodobacteraceae; Rhodobacter.
OX NCB1_TaxID=1063;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92283738; PubMed=1317839;
RA Meijer W.G., Tabita F.R.;
RT "Isolation and characterization of the nifusvW-zpon gene cluster from
RT Rhodobacter sphaeroides."
RL J. Bacteriol. 174:3855-3866(1992)
CC -1- SIMILARITY: BELONGS TO THE HESB/YADR/YFHF FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M86823; AAA26135.1; -
DR PIR; A41880; A41880.
DR InterPro; IPR000361; HESB_YADR_YFHF.
DR Pfam; PF01521; HESB-1ike; 1.
DR ProDom; PD002183; HESB_YADR_YFHF; 1.
DR TIGRFAMs; TIGR00049; TIGR00049; 1.
DR PROSITE; PS01152; HESB; 1.
KW Hypothetical protein.
SQ SEQUENCE 106 AA; 10780 MW; 5BAB3B3F7EADFC52 CRC64;

Query Match
Best Local Similarity 1.7%; Score 7; DB 1; Length 106;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 AAQAIRK 172
Db 7 AAQAIRK 13

RESULT 20
RL24_CHLPN
ID RL24_CHLPN STANDARD; PRT; 111 AA.
AC Q927R8; Q9J0G7;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 50S ribosomal protein L24.
GN RPLX OR RL24 OR CPN0636 OR CP0111.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCB1_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CWL029;
RC MEDLINE=99206606; PubMed=10192388;

```

RA Kalman S., Mitchell W., Marathe R., Iamuel C., Fan J., Hyman R.W.,
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
 RT "Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*";
 RU Nat. Genet. 21:385-389(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AR39;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Uterback T., Berry K., Bass S.,
 RA Liner K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwin W., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 RT "Genome sequences of *Chlamydia trachomatis* Mopn and *Chlamydia*
 RT *pneumoniae* AR39";
 RU Nucleic Acids Res. 28:1397-1406(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=J138;
 RX MEDLINE=20330349; PubMed=10871362;
 RA Shirai M., Hatakeyama H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
 RA Shiba T., Ishi K., Hattori M., Kuhara S., Nakazawa T.;
 RT "Comparison of whole genome sequences of *Chlamydia pneumoniae* J138
 RT from Japan and CwL029 from USA";
 RU Nucleic Acids Res. 28:2311-2314(2000).
 CC -1- FUNCTION: THIS PROTEIN IS FOUND IN THE RIBONUCLEOPROTEIN CORE AND
 CC IS INVOLVED IN THE EARLY ASSEMBLY OF THE 50S SUBUNIT. IT IS NOT
 CC INVOLVED IN THE FUNCTIONS OF THE MATURE 50S SUBUNIT (BY
 CC SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE L24P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE001647; AAD18775.1; -;
 DR EMBL; AE002173; AAF37994.1; -;
 DR EMBL; AF002547; BAA98843.1; -;
 DR PIR; A86570; A86570.
 DR PIR; F72054; F72054.
 DR TIGR; CP0111; -;
 DR InterPro; IPR005824; KOW
 DR InterPro; IPR006646; KOW_sub.
 DR InterPro; IPR003256; Ribosomal_L24.
 DR InterPro; IPR005825; Ribosomal_L24_26.
 DR Pfam; PF00467; KOW; 1.
 DR ProDom; PD001677; Ribosomal_L24; 1.
 DR SMART; SM00739; KOW; 1.
 DR TIGRfam; TIGR01079; rplX_bact; 1.
 DR PROSITE; PS01108; RIBOSOMAL_L24; 1.
 DR Ribosomal protein; Complete proteome.
 SQ SEQUENCE 111 AA; 12425 MW; 7BF9C0021CD7358F CRC64;
 Query Match 1.7%; Score 7; DB 1; Length 111;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 362 DKVVEG 368
 Db 32 DKVVEG 36
 RESULT 21
 RL20 CHLMU STANDARD; PRT; 123 AA.
 ID RL20 CHLMU
 AC Q9PL84;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE 50S ribosomal protein L20.
 GN RPLT OR TC0223.
 OS *Chlamydia muridarum*.
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83560;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Mopn / N199;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Uterback T., Berry K., Bass S.,
 RA Liner K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwin W., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 RT "Genome sequences of *Chlamydia trachomatis* Mopn and *Chlamydia*
 RT *pneumoniae* AR39";
 RU Nucleic Acids Res. 28:1397-1406(2000).
 CC -1- FUNCTION: This protein binds directly to 23s ribosomal RNA and is
 CC necessary for the in vitro assembly process of the 50s ribosomal
 CC subunit. It is not involved in the protein synthesizing functions
 CC of that subunit (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE L20P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE002289; AAF39095.1; -;
 DR PIR; E81727; E81727.
 DR TIGR; TC0223; -;
 DR HAMAP; MF_00382; -; 1.
 DR InterPro; IPR005813; L20.
 DR InterPro; IPR005812; L20_bact_orf.
 DR Pfam; PF00453; Ribosomal_L20; 1.
 DR PRINTS; PR00062; RIBOSOMAL_L20.
 DR ProDom; PD002389; L20; 1.
 DR TIGRfam; TIGR01032; rplT_bact; 1.
 DR PROSITE; PS00937; RIBOSOMAL_L20; 1.
 DR Ribosomal protein; rRNA-binding; Complete proteome.
 SQ SEQUENCE 123 AA; 13911 MW; 8B5373CE03640024 CRC64;
 Query Match 1.7%; Score 7; DB 1; Length 123;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 174 AGINLR 180
 Db 86 AGINLR 92
 RESULT 22
 SYGB NEIGO STANDARD; PRT; 126 AA.
 ID SYGB NEIGO
 AC Q50945;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Glycyl-tRNA synthetase beta chain (EC 6.1.1.14) (Glycine--tRNA ligase
 DE beta chain) (GLYRS) (Fragment).
 GN GLYS.
 OS Neisseria gonorrhoeae.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 OC Neisseriaceae; Neisseria.
 OX NCBI_TaxID=485;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 33084 / F62;
 RX MEDLINE=95053752; PubMed=7964493;
 RA Gotschlich E.C.;

```

RT "Genetic locus for the biosynthesis of the variable portion of
RT Neisseria gonorrhoeae lipooligosaccharide.";
RL J. Exp. Med. 180:2181-2190(1994).
CC -1- CATALYTIC ACTIVITY: ATP + glycine + tRNA(Gly) = AMP + diphosphate
CC + glycyl-tRNA(Gly).
CC -1- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS
CC (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U14554; AAA68008.1; -.
DR HAMAP; MF_00255; -.
DR InterPro; IPR006194; tRNA_synth_Gly.
DR PROSITE; PS50861; AA tRNA_LIGASE II GUYAB; 1.
KW Aminoacyl-tRNA synthetase; Protein Biosynthesis; Ligase; ATP-binding.
FT NON TER 1
SQ SEQUENCE 126 AA; 13343 MW; C3913AD1A6591CF2 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 126;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 AAALAAA 17
Db 13 AAALAAA 19

RESULT 23
TYRT STRGA STANDARD; PRT; 134 AA.
AC P55047;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Tyrosinase co-factor (URF402).
GN MELC1.
OS Streptomyces glaucescens.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1907;
RN SEQUENCE FROM N.A.
RP STRAIN=ATCC 13032 / ETH 22794 / GLA.0;
RX MEDLINE=88040431; PubMed=3118334;
RA Huber M., Huettler R., Lerch K.;
RL "The promoter of the Streptomyces glaucescens mel operon.";
Nucleic Acids Res. 15:8106-8106(1987).
-1- FUNCTION: THIS PROTEIN MAY FUNCTION TO DELIVER COPPER TO
TYROSINASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y00457; CA668512.1; -.
DR PIR; A26986; A26986.
DR InterPro; IPR006311; Tat.
DR TIGRPMs; TIGR01409; TAT_signal_seq; 1.
KW Melanin biosynthesis; Copper.
SQ SEQUENCE 134 AA; 13593 MW; 1B874145A7D1B9AE CRC64;

```

```

Query Match 1.7%; Score 7; DB 1; Length 134;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 AAALAAA 17
Db 13 AAALAAA 19

RESULT 24
RS12 HORVU
ID RS12 HORVU STANDARD; PRT; 143 AA.
AC Q9XHS0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 40S ribosomal protein S12.
GN RPS12.
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Triticeae; Hordeum.
OX NCBI_TaxID=4513;
RN SEQUENCE FROM N.A.
RP STRAIN=cv. Himalaya; TISSUE=Leaf;
RA Robertson M., Asami T.;
RT "An unique S12 protein sequence from barley, Hordeum vulgare L. cv.
RT Himalaya, with sequence relatedness to animal S12 proteins of 40S
RT small ribosome subunit.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE S12 FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF067732; AAD39838.1; -.
DR InterPro; IPR004038; Ribosomal_L7A.
DR InterPro; IPR000530; Ribosomal_S12e.
DR Pfam; PF01248; Ribosomal_L7Ae; 1.
DR PRINTS; PR00972; RIBSOMAL_S12E; 1.
DR PROSITE; PS01189; RIBSOMAL_S12E; 1.
KW Ribosomal protein.
SQ SEQUENCE 143 AA; 15295 MW; 6A7263198992A880 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 143;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 EAAPV 47
Db 4 EAAPV 10

RESULT 25
SYRB RHISN
ID SYRB RHISN STANDARD; PRT; 148 AA.
AC P55161;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Probable transcriptional regulator syrb.
GN SYRB OR Y4AN.
OS Rhizobium sp. (strain NGR234).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
OX NCBI_TaxID=394;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97305956; PubMed=9163424;
RA Freilberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA Perret X.,
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RL Nature 387:394-401(1997).
CC -1- FUNCTION: RESPONSIBLE FOR THE REPRESSION OF SYRM ACTIVITY
CC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE SYRB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, AE00065; AAB91611.1; -
CC Nucleotide sequence and predicted functions of the entire
CC SYRB2 OR RA0863 OR SWA1586.
CC Rhizobium meliloti (Sinorhizobium meliloti).
CC Plasmid pSymA (megaplasmid 1).
CC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
CC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
CC NCBI_TaxID=382;
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN=CMX1-105;
CC MEDLINE=99413305; PubMed=10485295;
CC Sharayova L.A., Yurgel S.N., Keller W., Simarov B.V., Puhler A.,
CC Becker A.;
RT "The eff-482 locus of Sinorhizobium meliloti CMX1-105 that influences
RT symbiotic effectiveness consists of three genes encoding an
RT endoglucanase, a transcriptional regulator and an adenylate cyclase.";
RL Mol. Gen. Genet. 261:1032-1044(1999).
CC [2]
CC SEQUENCE FROM N.A.
CC STRAIN=1021;
CC MEDLINE=21396509; PubMed=11481432;
CC Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
CC Barloy-Hubler F., Bowser L., Capela D., Galibert F., Guzy J.,
CC Gurjal M., Hong A., Hutzler L., Hyman R.W., Kahn D., Kahn M.L.,
CC Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
CC Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
RT "Nucleotide sequence and predicted functions of the entire
RT Sinorhizobium meliloti pSymA megaplasmid.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
CC -1- FUNCTION: SEEMS TO AFFECT THE TRANSCRIPTION OF CYA3. MAY BE
CC NEGATIVELY AUTOREGULATED.
CC -1- SIMILARITY: BELONGS TO THE SYRB FAMILY.
CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO TWO
CC FRAMESHIFTS IN POSITION 124 AND 137.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, AJ225896; CAB38102.1; ALT_FRAME.
CC EMBL, AE007223; AAK65521.1; -
CC PIR, G95369; G95369.
CC InterPro: IPR002514; Transposase_8.
CC Pfam: PF01527; Transposase_8; 1.
CC Nucleotide sequence and predicted functions of the entire
CC complete proteome.
CC CONFLICT 17 N -> H (IN REF. 1).
CC SEQUENCE 151 AA; 17028 MW; 215627A0FADBBAB8 CRC64;
SQ
Query Match 1.7%; Score 7; DB 1; Length 151;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 170 AIKSAGI 176
DB 81 AIKSAGI 87
RESULT 27
SYB3_RHIME STANDARD; PRT; 151 AA.
AC P58346;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable transcriptional regulator syrb3.
GN SYRB3 OR RA0438 OR SWA0806.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OG Plasmid pSymA (megaplasmid 1).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
CC NCBI_TaxID=382;
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN=1021;
CC MEDLINE=21396509; PubMed=11481432;
CC Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
CC Barloy-Hubler F., Bowser L., Capela D., Galibert F., Guzy J.,
CC Gurjal M., Hong A., Hutzler L., Hyman R.W., Kahn D., Kahn M.L.,
CC Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
CC Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
RT "Nucleotide sequence and predicted functions of the entire
RT Sinorhizobium meliloti pSymA megaplasmid.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
CC -1- SIMILARITY: BELONGS TO THE SYRB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, AE007234; AAK5096.1; -
CC PIR, F95316; F95316.
CC Nucleotide sequence and predicted functions of the entire
CC complete proteome.
CC CONFLICT 151 AA; 16795 MW; 9CC59F4B6123B28 CRC64;
SQ
Query Match 1.7%; Score 7; DB 1; Length 151;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```


OY 170 AIKSAGI 176
 DB 78 AIKSAGI 84

RESULT 28

PSBU CYACA STANDARD; PRT; 154 AA.
 ID PSBU CYACA 092055; 092780;
 AC 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Photosystem II 12 kDa extrinsic protein, chloroplast precursor (PS II complex 12 kDa extrinsic protein) (PSII-U).
 OS PSBU.
 OS Cyanidium caldarium.
 OC Eukaryota; Rhodophyta; Bangiophyceae; Porphyridiales; Porphyridiaceae;
 OC Cyanidium.
 OX NCBI_TaxID=2771;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99310600; PubMed=10381374;
 RA Ohta H., Okumura A., Okuyama S., Akiyama A., Iwai M., Yoshinara S.,
 RA Shen J.-R., Kamo M., Enami I.,
 RT "Cloning, expression of the psbU gene, and functional studies of the recombinant 12-kDa protein of photosystem II from a red alga
 RT Cyanidium caldarium.";
 RL Biochem. Biophys. Res. Commun. 260:245-250 (1999).
 RN [2]
 RP SEQUENCE OF 62-112.
 RX MEDLINE=96124997; PubMed=8534673;
 RA Enami I., Murayama H., Ohta H., Kamo M., Nakazato K., Shen J.-R.;
 RT "Isolation and characterization of a Photosystem II complex from the red alga Cyanidium caldarium: association of cytochrome c-550 and a 12 kDa protein with the complex.";
 RL Biochim. Biophys. Acta 1232:208-216 (1995).
 CC -1- FUNCTION: Constitutively stabilizes and protects the oxygen-evolving complex of photosystem II against heat-induced
 CC inactivation (By similarity).
 CC -1- SUBCELLULAR LOCATION: Loosely bound to the inner surface of the chloroplast thylakoid membrane.
 CC -1- SIMILARITY: BELONGS TO THE PSBU FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, AB023805; BAA75396.1; AUT_INIT.
 DR TIGRFAAS; TIGR01409; TAT_signal_seq; 1.
 KW Photosynthesis; Photosystem II; Thylakoid; Membrane; Transit peptide;
 KW Chloroplast.
 FT TRANSIT 1 ? CHLOROPLAST (POTENTIAL).
 FT TRANSIT 2 ? THYLAKOID.
 FT CHAIN 62 154 PHOTOSYSTEM II 12 KDA EXTRINSIC PROTEIN.
 FT CONFLICT 80 80 N -> D (IN REF. 2).
 FT CONFLICT 90 90 F -> I (IN REF. 2).
 SQ SEQUENCE 154 AA; 16715 MW; EF28D01ABEC09777 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 154;
 Best local Similarity 100.0%; Pred. No. 46;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 ALAAVA 19
 DB 44 ALAAVA 50

RESULT 29

CYP2 ARATH STANDARD; PRT; 169 AA.
 ID CYP2 ARATH P35627;
 AC 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Peptidyl-prolyl cis-trans isomerase (EC 5.2.1.8) (PPIase) (Notamase)
 DE (Cyclophilin) (Cyclosporin A-binding protein).
 OS Arabidopsis thaliana (Mouse-ear cress).
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eucosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Landsberg erecta; TISSUE=leaf.
 RX MEDLINE=92322989; PubMed=1623198;
 RA Bartling D., Heese A., Weller E.W.;
 RT "Nucleotide sequence of a cDNA encoding an Arabidopsis
 RT cyclophilin-like protein.";
 RL Plant Mol. Biol. 19:529-530 (1992).
 CC -1- FUNCTION: PPIases accelerate the folding of proteins. It catalyzes the cis-trans isomerization of proline imidic peptide bonds in oligopeptides.
 CC -1- CATALYTIC ACTIVITY: Peptidylproline (omega=180) = peptidylproline (omega=0).
 CC -1- ENZYME REGULATION: BINDS CYCLOSPORIN A (CSA). CSA MEDIATES SOME OF ITS EFFECTS VIA AN INHIBITORY ACTION ON PPIASE.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -1- SIMILARITY: BELONGS TO THE CYCLOPHILIN-TYPE PPIASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, X63616; CAA45161.1; -.
 DR PIR, S22496; S22496.
 DR HSSP; P05092; 2CPL.
 DR InterPro; IPR002130; CSA_PPIase.
 DR Pfam; PF00160; pro_isomerase; 1.
 DR PRINTS; PR00153; CSAPISMASE.
 DR PROSITE; PS00170; CSA_PPIASE_1; 1.
 DR PROSITE; PS50072; CSA_PPIASE_2; 1.
 KW Cyclosporin; Isomerase; Rotamase; Multigene family.
 SQ SEQUENCE 169 AA; 18161 MW; B6077FC139864931 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 169;
 Best local Similarity 100.0%; Pred. No. 50;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 34 GGOPAGR 40
 DB 11 GGOPAGR 17

RESULT 30
 YSF0 STROO STANDARD; PRT; 169 AA.
 ID YSF0 STROO 092863;
 AC 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical protein SC06450.
 GN SC06450 OR SC985.17.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycinae; Streptomycetaceae; Streptomycetes.
 OX NCBI_TaxID=1902;
 RN [1]

```

RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kisser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleser T., Lakte L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Watzorek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RU Nature 417:141-147(2002).
CC -1- SIMILARITY: BELONGS TO THE UPF0189 FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AL035206; CAA22759.1; -
DR PIR; T35937; T35937.
DR HAMAP; MF_01205; -; 1.
DR InterPro; IPR002589; A1pp.
DR Pfam; PF01601; A1pp; 1.
DR SMART; SM00506; A1pp; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 169 AA; 17767 MW; 86965DB33C6E4740 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 169;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 68 ESLRTAD 74
Db 103 ESLRTAD 109

RESULT 31
CYPH CATRO STANDARD; PRT; 172 AA.
AC Q39613;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Peptidyl-prolyl cis-trans isomerase (EC 5.2.1.8) (PPIase) (Rotamase)
DE (Cyclophilin) (Cyclosporin A-binding protein).
GN PCK1.
OS Catharanthus roseus (Rosa periwinkle) (Madagascar periwinkle).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamiales; Gentianales; Apocynaceae; Rauvolfioideae; Vincaceae;
OC Catharanthus.
OC NCBI_TaxID=4058;
OK [1]
RN SEQUENCE FROM N.A.
RA Claessens M., Maeroufi H., Andreu F., Chenieux J.-C., Rideau M.,
RA Hamdi S.;
RT "Isolation of a full-length cDNA encoding a cytosolic cyclophilin from
RT periwinkle (Catharanthus roseus).";
RL (In) Plant Gene Register PGR95-100.
CC -1- FUNCTION: PPIase accelerate the folding of proteins. It catalyzes
CC the cis-trans isomerization of proline imidic peptide bonds in
CC oligopeptides.
CC -1- CATALYTIC ACTIVITY: Peptidylproline (omega=180) = peptidylproline
CC (omega=0).
CC -1- ENZYME REGULATION: BINDS CYCLOSPORIN A (CSA). CSA MEDIATES SOME
CC OF ITS EFFECTS VIA AN INHIBITORY ACTION ON PPIASE.

```

```

CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE CYCLOPHILIN-TYPE PPIASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X85185; CAA59468.1; -
DR PIR; T10056; T10056.
DR HSSP; P05092; 1CWL.
DR InterPro; IPR002130; CSA_PPIase.
DR Pfam; PF00160; pro_isomerase; 1.
DR PRINTS; PR00153; CSA_PPIASE.
DR PROSITE; PS00170; CSA_PPIASE_1; 1.
DR PROSITE; PS50072; CSA_PPIASE_2; 1.
KW Cyclosporin; Isomerase; Rotamase; Multigene family.
SQ SEQUENCE 172 AA; 18285 MW; EA6EC51886A50A81 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 172;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 34 GGQAPGR 40
Db 13 GGQAPGR 19

RESULT 32
IF3 ANASP STANDARD; PRT; 177 AA.
ID IF3 ANASP
AC Q8WTE3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Translation initiation factor IF-3.
GN INFC OR ALL4623.
OS Anabaena sp. (Strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriyuchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Rep. 8:205-213(2001).
CC -1- FUNCTION: IF-3 binds to the 30S ribosomal subunit and shifts the
CC equilibrium between 70S ribosomes and their 50S and 30S subunits in
CC favor of the free subunits, thus enhancing the availability of 30S
CC subunits on which protein synthesis initiation begins.
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to the IF-3 family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AP003597; BAB76322.1; ALT_INIT.
DR HAMAP; MF_00080; -; 1.
DR InterPro; IPR001288; IF3.

```

DR Pfam; PF00707; IF3-C; 1.
 DR Pfam; PF05198; IF3-N; 1.
 DR ProDom; PD002880; IF3; 1.
 DR TIGRFAMs; TIGR00168; infc; 1.
 DR PROSITE; PS00938; IF3; 1.
 KW Initiation factor; Protein biosynthesis; Complete proteome.
 SQ SEQUENCE 177 AA; 20750 MW; 9C1585FF62590EFP CRC64;

Query Match 1.7%; Score 7; DB 1; Length 177;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 358 LKDGDKV 364
 DB 118 LKDGDKV 124

RESULT 33
 ID IF3_THETN STANDARD; PRT; 180 AA.
 AC Q8R9C2;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Translation initiation factor IF-3.
 GN INFC OR TTE1693.
 OS Thermomicrobacter tengcongensis.
 OC Bacteria; Firmicutes; Clostridia; Thermomicrobacteriales;
 OC Thermomicrobacteriaceae; Thermomicrobacter.
 OK NCBI_TaxID=119072;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MB4 / JCM 11007;
 RX MEDLINE=21992816; PubMed=11997336;
 RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
 Chen Y., Xue Y., Xu Y., Lai X., Huang L.,
 Tan H., Chen R., Wang J., Yu J., Yang H.;
 RT "A complete sequence of T. tengcongensis genome.";
 RL Genome Res. 12:689-700(2002).
 CC -1- FUNCTION: IF-3 binds to the 30S ribosomal subunit and shifts the
 equilibrium between 70S ribosomes and their 50S and 30S subunits in
 favor of the free subunits, thus enhancing the availability of 30S
 subunits on which protein synthesis initiation begins.
 CC -1- SUBUNIT: Monomer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: Belongs to the IF-3 family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE013124; AM24894.1; -.
 DR HAMAP; MF_00080; -; 1.
 DR InterPro; IPR001288; IF3.
 DR Pfam; PF00707; IF3-C; 1.
 DR Pfam; PF05198; IF3-N; 1.
 DR ProDom; PD002880; IF3; 1.
 DR TIGRFAMs; TIGR00168; infc; 1.
 DR PROSITE; PS00938; IF3; 1.
 KW Initiation factor; Protein biosynthesis; Complete proteome.
 SQ SEQUENCE 180 AA; 21061 MW; 83983538F8FCBB CRC64;

Query Match 1.7%; Score 7; DB 1; Length 180;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 358 LKDGDKV 364
 DB 120 LKDGDKV 126

RESULT 34
 ID KRUC_SHEEP STANDARD; PRT; 182 AA.
 AC P26372;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Keratin, ultra high-sulfur matrix protein (UHS keratin).
 OS Ovis aries (sheep).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Caprinae; Ovis.
 OK NCBI_TaxID=9940;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Follicle;
 RX MEDLINE=91115951; PubMed=1703541;
 RA McKinnon P.J., Powell B.C., Rogers G.E.;
 RT "Structure and expression of genes for a class of cysteine-rich
 proteins of the cuticle layers of differentiating wool and hair
 follicles.";
 RL J. Cell Biol. 111:2587-2600(1990).
 CC -1- FUNCTION: THE KERATIN PRODUCTS OF MAMMALIAN EPIDERMAL DERIVATIVES
 SUCH AS WOOL AND HAIR CONSIST OF MICROFIBRILS EMBEDDED IN A RIGID
 MATRIX OF OTHER PROTEINS. THE MATRIX PROTEINS INCLUDE THE HIGH-
 SULFUR AND HIGH-TYROSINE KERATINS, HAVING MOLECULAR WEIGHTS OF
 6-20 KDa, WHEREAS THE MICROFIBRILS CONTAIN THE LARGER, LOW-SULFUR
 KERATINS (40-56 kDa).
 CC -1- TISSUE SPECIFICITY: CUTICLE LAYERS OF DIFFERENTIATING
 WOOL FOLLICLES.
 CC -1- DEVELOPMENTAL STAGE: AT A LATE STAGE OF FIBER DIFFERENTIATION.
 CC -1- DOMAIN: MAINLY COMPOSED OF CYS-RICH (CR), GLY-RICH (GR) AND SER-
 RICH (SR) REPEATS.
 CC -1- SIMILARITY: BELONGS TO THE UHS KERATIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X55294; CA339006.1; -.
 DR PIR; A36586; A36586.
 DR HSSP; P02876; 9WGA.
 KW Keratin; Repeat; Multigene family.
 SQ SEQUENCE 182 AA; 16101 MW; 9BDD4901FCB13295 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 182;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 23 SSCGKGG 29
 DB 48 SSCGKGG 54

RESULT 35
 ID RBPA_MYCTU STANDARD; PRT; 183 AA.
 AC P1614;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable ribosome-binding factor A.
 GN RBPA OR RV2838C OR MT2904 OR MTCY1687.04.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OK NCBI_TaxID=1773;

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigleier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jorgels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulten J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
CC -1- FUNCTION: Associates with free 30S ribosomal subunits (but not
CC with 30S subunits that are part of 70S ribosomes or polysomes).
CC Essential for efficient processing of 16S rRNA. May interact with
CC the 5' terminal helix region of 16S rRNA (by similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- SIMILARITY: BELONGS TO THE RBFA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; 281331; CAB03671.1; -
CC EMBL; AE007115; AAK47230.1; -
CC DR PIR; A70694; A70694.
CC DR TIGR; MT2904; -.
CC DR TubercuList; RV2838c; -.
CC DR HAMAP; MF_00003; -.
CC DR InterPro; IPR000238; Rib_bind_facta.
CC DR Pfam; PF02033; RBFA; 1.
CC DR ProDom; PD007327; Rib_bind_facta; 1.
CC DR TIGRFAMs; TIGR00082; Ibfa; 1.
CC DR PROSITE; PS01319; RBFA; 1.
CC DR RNA processing; Complete proteome.
CC SQ SEQUENCE 183 AA; 18998 MW; DCEFLA043CB34F92 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 183;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 ADADLAR 135
DB 119 ADADLAR 125

RESULT 36
ID IF3_RICPR STANDARD; PRT; 185 AA.
AC 092D19;
RX 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Translation initiation factor IF-3.
GN INFC OR RP531.
RT

```

```

OS 05 Rickettsia prowazekii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiaseae; Rickettsia.
OX NCBI_TaxId=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Madrid E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sichelitz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria."
RL Nature 396:133-140 (1998).
CC -1- FUNCTION: IF-3 binds to the 30S ribosomal subunit and shifts the
CC equilibrium between 70S ribosomes and their 50S and 30S subunits in
CC favor of the free subunits, thus enhancing the availability of 30S
CC subunits on which protein synthesis initiation begins.
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to the IF-3 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ235272; CAA14980.1; -
CC DR PIR; B71657; B71657.
CC DR HSSP; P03000; ITIF.
CC DR HAMAP; MF_00080; -.
CC DR InterPro; IPR001288; IF3.
CC DR Pfam; PF00707; IF3_C; 1.
CC DR Pfam; PF05198; IF3_N; 1.
CC DR ProDom; PD002880; IF3; 1.
CC DR TIGRFAMs; TIGR00168; Infc; 1.
CC DR PROSITE; PS00938; IF3; 1.
CC DR Initiation factor; Protein biosynthesis; Complete proteome.
CC SQ SEQUENCE 185 AA; 21605 MW; EB483C963B6581BC CRC64;

Query Match 1.7%; Score 7; DB 1; Length 185;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 358 LKGDGV 364
DB 124 LKGDGV 130

RESULT 37
ID IF3_RICCN STANDARD; PRT; 186 AA.
AC 092HK6;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Translation initiation factor IF-3.
GN INFC OR RC0765.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiaseae; Rickettsia.
OX NCBI_TaxId=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii."

```

```

RL Science 293:2093-2098(2001).
CC -1- FUNCTION: IF-3 binds to the 30S ribosomal subunit and shifts the
CC equilibrium between 70S ribosomes and their 50S and 30S subunits in
CC favor of the free subunits, thus enhancing the availability of 30S
CC subunits on which protein synthesis initiation begins.
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to the IF-3 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE008633; AAL03303.1; ALT_INIT.
DR HAMAP; MF_00080; -, 1.
DR InterPro; IPR001288; IF3.
DR Pfam; PF00707; IF3_C; 1.
DR Pfam; PF05198; IF3_N; 1.
DR ProDom; PD002880; IF3; 1.
DR TIGRfam; TIGR00168; INF3; 1.
DR PROSITE; PS00938; IF3; 1.
DR Initiation factor; Protein biosynthesis; Complete proteome.
DR KW
DR SEQUENCE 186 AA; 21665 MW; B42202AD4B293043 CRC64;
SQ
Query Match 1.7%; Score 7; DB 1; Length 186;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 358 LKDGDKV 364
DB 124 LKDGDKV 130
-----
RESULT 38
Y418_VIBCH STANDARD; PRT; 187 AA.
ID Y418_VIBCH
AC 09KUT7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Maf-like protein VCO418.
GN VCO418.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406633; PubMed=10952301;
RA Heddelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Baas S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Ueberback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RA "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RA cholerae";
RT Nature 406:477-483(2000).
RL -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- SIMILARITY: BELONGS TO THE MAF FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE004129; AAF93591.1; ALT_INIT.
DR HSSP; 002169; 1EX2.
DR TIGR; VC0418; -.
DR HAMAP; MF_00528; -, 1.
DR InterPro; IPR003697; Maf.
DR Pfam; PF02545; Maf; 1.
DR TIGRfam; TIGR00172; maf; 1.
DR KW
DR Complete proteome.
DR ACT SITE 35
DR SEQUENCE 187 AA; 20495 MW; DF7DA9F27667C0F7 CRC64;
SQ
Query Match 1.7%; Score 7; DB 1; Length 187;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 139 LVAAEAV 145
DB 61 LVAAEAV 67
-----
RESULT 39
PAT_ALCEFA STANDARD; PRT; 197 AA.
ID PAT_ALCEFA
AC P31668;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Phosphinothricin N-acetyltransferase (EC 2.3.1.-) (PPT N-
DE acetyltransferase) (Phosphinothricin-resistance protein).
GN PAT.
OS Alcaligenes faecalis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Alcaligenes.
OX NCBI_TaxID=511;
RN [1]
RP SEQUENCE FROM N.A.
RC Brauer D., Bartsch K., Donn G.;
RT "Gene resistant to phosphinothricin";
RL Patent number EP0290986, 17-NOV-1988.
CC -1- FUNCTION: THIS ENZYME IS AN EFFECTOR OF PHOSPHINOTHRICIN
CC TRIPEPTIDE (PPT OR BIALAPHOS) RESISTANCE. INACTIVATES PPT BY
CC TRANSFER OF AN ACETYL GROUP.
CC -1- SIMILARITY: BELONGS TO THE ACETYLTRANSFERASE FAMILY. PAT/BAR
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; A01505; -; NOT ANNOTATED - CDS.
DR EMBL; A01504; CAA00175.1; -.
DR InterPro; IPR000182; GCN5acetyltransf.
DR Pfam; PF00583; Acetyltransf; 1.
DR Transferrase; Acyltransferase; Herbicide resistance;
DR Antibiotic resistance.
DR KW
DR SEQUENCE 197 AA; 21213 MW; ACA6E3B51AA5FE3 CRC64;
SQ
Query Match 1.7%; Score 7; DB 1; Length 197;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 202 LNAAGDT 208
DB 188 LNAAGDT 194
-----
RESULT 40

```

COAE BRUME
 ID COAE BRUME STANDARD; PRT; 200 AA.
 AC 08YE21;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Dephospho-CoA kinase (EC 2.7.1.24) (Dephosphocoenzyme A kinase).
 GN COAE OR BMEI2057 OR BR2070.
 OS Brucella melitensis, and
 OS Brucella suis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Brucellaceae; Brucella.
 OK NCBI_TaxID=29459, 29461;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=B. melitensis; STRAIN=16M / ATCC 23456 / Biotype 1;
 RX MEDLINE=20020109; PubMed=11756588;
 RA DelVecchio V.G., Kapatal V., Redkar R.J., Patra G., Mujer C., Los T.,
 RA Ivanova N., Anderson I., Bhattacharya A., Iykidis A., Reznik G.,
 RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldsman E.,
 RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
 RA Haselkorn R., Kyriades N., Overbeek R.;
 RA "The genome sequence of the facultative intracellular pathogen
 RT Brucella melitensis.";
 RT Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=B. suis; STRAIN=1330 / Biovar 1;
 RX MEDLINE=22247741; PubMed=12271122;
 RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
 RA Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,
 RA Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
 RA Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,
 RA Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,
 RA Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M., Fraser C.M.;
 RT "The Brucella suis genome reveals fundamental similarities between
 RT animal and plant pathogens and symbionts.";
 RT Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).
 CC -!- FUNCTION: Catalyzes the phosphorylation of the 3'-hydroxyl group
 CC of dephosphocoenzyme A to form coenzyme A (By similarity).
 CC -!- CATALYTIC ACTIVITY: ATP + dephospho-CoA = ADP + CoA.
 CC -!- PATHWAY: Coenzyme A (CoA) biosynthesis; fitch (last) step.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: Belongs to the coa family.
 CC
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC
 CC EMBL; AE009638; AAL53238.1; -;
 CC EMBL; AE014494; AAN30960.1; -;
 CC PIR; AC3509; AC3509.
 CC TIGR; BR2070; -;
 CC HAMAP; MF_00376; -; 1.
 CC InterPro; IPR001977; Depp_CoAkinase.
 CC Pfam; PF01121; Coae; 1.
 CC ProDom; PD003329; Depp_CoAkinase; 1.
 CC TIGRFAMs; TIGR00152; TIGR00152; 1.
 CC PROSITE; PS01294; COAE; FALSE NEG.
 CC Transferase; kinase; ATP-binding; Coenzyme A biosynthesis;
 CC Complete proteome.
 CC NP_BIND 8 15 ATP (POTENTIAL).
 CC SEQUENCE 200 AA; 21441 MW; B2CB67FF86C8430D CRC64;

Db 180 LRRQIAE 186

Search completed: September 8, 2003, 14:09:05
 UOB time : 32 secs

Query Match 1.7%; Score 7; DB 1; Length 200;
 Best Local Similarity 100.0%; Pred. No. 57;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 233 LRRQIAE 239

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2003, 14:06:27 ; Search time 21 Seconds
(without alignments)
1886.738 Million cell updates/sec

Title: US-09-889-756A-2

Perfect score: 412
Sequence: 1 MAFYAFKAMRAALAAVAL.....AAPQGVQTASEAKTASEAE 412

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : PIR_76:*

1: pir1:*\n2: pir2:*\n3: pir3:*\n4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	171	41.5	412	2	H81825
2	133	32.3	412	2	F81051
3	100	24.3	271	2	S42418
4	11	2.7	373	2	E85757
5	11	2.7	373	2	G90861
6	11	2.7	388	2	E87349
7	9	2.2	444	2	AH0346
8	8	1.9	107	2	C81232
9	8	1.9	111	2	E42645
10	8	1.9	111	2	D81664
11	8	1.9	122	2	D70730
12	8	1.9	129	2	C70022
13	8	1.9	171	2	D90703
14	8	1.9	171	2	G85553
15	8	1.9	171	2	B64785
16	8	1.9	181	2	A70673
17	8	1.9	183	2	A72499
18	8	1.9	183	2	G90849
19	8	1.9	183	2	F85707
20	8	1.9	290	2	T36757
21	8	1.9	296	2	C95961
22	8	1.9	306	2	H86740
23	8	1.9	348	2	C70415
24	8	1.9	408	2	AC0961
25	8	1.9	416	2	B65171
26	8	1.9	416	2	B81207
27	8	1.9	416	2	G86053
28	8	1.9	434	1	WMBE83
29	8	1.9	478	1	DPHDM2

30	8	1.9	480	2	A46702	methionyl aminope
31	8	1.9	489	2	T22088	hypothetical prote
32	8	1.9	716	2	AB1070	probable carbon st
33	8	1.9	748	2	T35818	probable integral
34	8	1.9	871	2	AH0172	membrane alanyl am
35	8	1.9	891	1	DEEC	acetaldehyde dehyd
36	8	1.9	891	2	D85704	hypothetical prote
37	8	1.9	891	2	F90846	acetaldehyde dehyd
38	8	1.9	891	2	AC0265	alcohol dehydrogen
39	8	1.9	892	2	AE0650	alcohol dehydrogen
40	8	1.9	979	2	T01566	hypothetical prote
41	8	1.9	984	2	T44496	cellulose 1,4-beta
42	8	1.9	1121	2	T02764	myosin-I binding p
43	8	1.9	1291	2	T13389	hypothetical prote
44	8	1.9	1558	2	C89314	protein C37C.6a l
45	8	1.9	1882	2	T00069	hypothetical prote
46	8	1.9	2167	2	T34395	hypothetical prote
47	8	1.7	13	2	A38929	glutathione peroxi
48	8	1.7	79	2	S61446	cellulase (EC 3.2.
49	7	1.7	91	1	RGE8BD	mercuric resistanc
50	7	1.7	91	1	S09524	mercuric resistanc
51	7	1.7	97	2	T28947	hypothetical prote
52	7	1.7	98	2	T28629	hypothetical prote
53	7	1.7	106	2	S07520	Y4C8 protein - Rhl
54	7	1.7	106	2	A41880	host specificity p
55	7	1.7	111	2	F72054	nitrogen fixation
56	7	1.7	111	2	A86570	ribosomal protein
57	7	1.7	123	2	B81727	L24 ribosomal prot
58	7	1.7	131	2	AF3339	ribosomal protein
59	7	1.7	132	2	A93339	precorrin-3b C17-m
60	7	1.7	134	2	I39004	cyclin-dependent k
61	7	1.7	135	2	A26986	hypothetical prote
62	7	1.7	135	2	F87515	hypothetical prote
63	7	1.7	145	2	AD2920	hypothetical prote
64	7	1.7	145	2	B7664	cofamin biosynth
65	7	1.7	147	2	H70514	hypothetical prote
66	7	1.7	151	2	F95316	probable SyxB-like
67	7	1.7	152	2	G95369	SVB2 transcrip
68	7	1.7	152	2	A87264	hypothetical prote
69	7	1.7	154	2	AE0445	biotin carboxyl ca
70	7	1.7	157	2	E97795	translation initia
71	7	1.7	157	2	C71060	hypothetical prote
72	7	1.7	158	2	A12718	conserved hypotet
73	7	1.7	158	2	E97500	hypothetical prote
74	7	1.7	163	2	T15985	hypothetical prote
75	7	1.7	168	2	AE0138	peptidoglycan-asso
76	7	1.7	169	2	S22496	peptidylprolyl iso
77	7	1.7	172	2	T35937	hypothetical prote
78	7	1.7	172	2	S54833	peptidylprolyl iso
79	7	1.7	172	2	T10056	peptidylprolyl iso
80	7	1.7	180	2	F75607	atrenate reductase
81	7	1.7	181	2	H69157	hypothetical prote
82	7	1.7	182	2	A36686	ultra-high-sulfur
83	7	1.7	183	2	A70694	probable rbfA - My
84	7	1.7	185	2	B71657	translation initia
85	7	1.7	187	2	B65100	hypothetical 19.4
86	7	1.7	191	2	T46412	keratin KAP5.4 - s
87	7	1.7	197	2	T46413	keratin KAP5.5 - s
88	7	1.7	200	2	AC3509	dephospho-CoA kina
89	7	1.7	202	2	B87342	hypothetical prote
90	7	1.7	203	2	T36403	probable tetr-fami
91	7	1.7	204	2	AG2383	translation initia
92	7	1.7	205	1	OPBOE	glutathione peroxi
93	7	1.7	205	2	C82325	mal protein VC0418
94	7	1.7	206	2	T36643	probable integral
95	7	1.7	209	2	D81270	hypothetical prote
96	7	1.7	209	2	G87569	conserved hypotet
97	7	1.7	212	2	AD2356	AhpC/GSA family pr
98	7	1.7	212	2	C75049	hypothetical prote
99	7	1.7	213	2	F84532	hypothetical prote
100	7	1.7	214	2	A97624	hypothetical prote
101	7	1.7	214	2	AH2846	hydrolase [importe
102	7	1.7	217	2	T06591	cellulase (EC 3.2.
			223	2	T35908	probable membrane

103	7	1.7	226	2	T06378	oleosin p24, isofo
104	7	1.7	230	2	E87578	hypothetical prote
105	7	1.7	231	2	D69983	conserved hypothet
106	7	1.7	232	2	C71908	3-oxoacid CoA tran
107	7	1.7	233	2	T43285	probable progester
108	7	1.7	240	2	AE3612	multidrug resistan
109	7	1.7	242	2	D81438	dihydrodipicolinat
110	7	1.7	242	2	B82815	conserved hypothet
111	7	1.7	243	2	T05307	hypothetical prote
112	7	1.7	244	2	AB0195	3-oxoacyl-l-acyl-ca
113	7	1.7	249	2	E82917	methionine aminope
114	7	1.7	251	2	S39202	cellulase (EC 3.2.
115	7	1.7	260	2	C69130	shikimate 5-dehydr
116	7	1.7	267	2	E75471	transcription regu
117	7	1.7	270	2	JN0767	homobox protein H
118	7	1.7	271	2	S30230	homeotic protein H
119	7	1.7	273	2	AB2811	conserved hypothet
120	7	1.7	273	2	E97589	hypothetical prote
121	7	1.7	274	2	S67444	probable 5'-AMP-ac
122	7	1.7	275	2	C87153	acyl-[ACP] desatur
123	7	1.7	277	2	F84336	pyrroline-5-carbox
124	7	1.7	278	1	G69964	ribose phosphate p
125	7	1.7	281	2	A84371	hypothetical prote
126	7	1.7	283	2	B82631	conserved hypothet
127	7	1.7	286	2	H83101	hypothetical prote
128	7	1.7	286	2	E69516	hypothetical prote
129	7	1.7	288	2	T45715	hypothetical prote
130	7	1.7	293	2	S77419	hypothetical prote
131	7	1.7	295	2	AC1268	cell-shape determi
132	7	1.7	295	2	AE1630	cell-shape determi
133	7	1.7	302	2	A96841	hypothetical prote
134	7	1.7	308	2	F82509	conserved hypothet
135	7	1.7	312	2	S18387	heme oxygenase - r
136	7	1.7	315	2	E81937	probable transmemb
137	7	1.7	315	2	B81168	transporter NMB070
138	7	1.7	317	2	AB0450	trehalase operon r
139	7	1.7	319	2	S61430	cellulase (EC 3.2.
140	7	1.7	324	2	S61447	cellulase (EC 3.2.
141	7	1.7	324	2	D82845	copper resistance
142	7	1.7	324	2	T23876	hypothetical prote
143	7	1.7	326	2	B40141	mitochondrial solu
144	7	1.7	330	2	F81656	conserved hypothet
145	7	1.7	332	2	S76622	hypothetical prote
146	7	1.7	334	2	T03520	probable nicotinat
147	7	1.7	336	2	F95925	probable cell-wall
148	7	1.7	349	2	A40141	microchondrial solu
149	7	1.7	350	2	T36611	probable integral
150	7	1.7	354	2	T04779	hypothetical prote
151	7	1.7	355	2	JC1249	peroxidase (EC 1.1
152	7	1.7	355	2	F90391	conserved hypothet
153	7	1.7	357	2	J01647	SH1 protein - hum
154	7	1.7	358	2	C82281	ferric vibriobacti
155	7	1.7	361	2	T48297	protein phosphatas
156	7	1.7	363	2	T44150	hypothetical prote
157	7	1.7	364	2	S34355	peroxidase (EC 1.1
158	7	1.7	367	2	T34884	probable integral
159	7	1.7	369	2	E81379	probable membrane
160	7	1.7	369	2	B87645	HLV family secret
161	7	1.7	372	2	A36693	lignin peroxidase
162	7	1.7	373	2	B70673	probable ddla - My
163	7	1.7	374	2	T04778	hypothetical prote
164	7	1.7	376	2	C95878	probable dehydrog
165	7	1.7	380	2	AE3324	UDP-N-acetylglucos
166	7	1.7	380	2	D70516	hypothetical prote
167	7	1.7	383	2	T45819	villin 3 homolog F
168	7	1.7	385	2	T36688	probable serine pr
169	7	1.7	386	2	S52981	lycogene cyclase -
170	7	1.7	393	2	D81139	2-oxoglutarate deh
171	7	1.7	393	2	T35894	hypothetical prote
172	7	1.7	395	2	T35481	hypothetical prote
173	7	1.7	396	2	A70581	hypothetical prote
174	7	1.7	396	2	E83393	RND multidrug effl
175	7	1.7	398	2	F69266	3-ketoacyl-CoA thl
176	7	1.7	402	2	T34715	probable ornithine
177	7	1.7	403	2	A81882	probable dihydroli
178	7	1.7	404	2	D70977	hypothetical prote
179	7	1.7	405	2	A36896	tetracycline resis
180	7	1.7	405	2	A32812	repa protein - Agr
181	7	1.7	406	2	G72538	hypothetical prote
182	7	1.7	409	2	T35118	probable secreted
183	7	1.7	417	2	S51961	FUN50 protein - ye
184	7	1.7	417	2	G82968	serine hydroxymeth
185	7	1.7	417	2	A75563	2-oxoglutarate deh
186	7	1.7	417	2	B83487	probable oxidoredu
187	7	1.7	418	2	C83341	serine hydroxymeth
188	7	1.7	418	2	E72776	hypothetical prote
189	7	1.7	420	2	H84360	hypothetical prote
190	7	1.7	433	2	T43961	hypothetical prote
191	7	1.7	435	2	T36780	hypothetical prote
192	7	1.7	436	2	F91127	hypothetical prote
193	7	1.7	436	2	B85972	hypothetical prote
194	7	1.7	441	2	A75619	cobyrinic acid a,c
195	7	1.7	441	2	T48960	vacuolar H(+)-ATP
196	7	1.7	447	2	B82490	probable Na+/H+ an
197	7	1.7	447	2	G84001	glycine dehydrogen
198	7	1.7	458	2	A48392	alpha 2C4 adrenoce
199	7	1.7	458	2	A40392	alpha-2-adrenoergi
200	7	1.7	458	2	A37869	alpha-2B-adrenoergi
201	7	1.7	458	2	T49480	alpha-2 adrenoergi
202	7	1.7	458	2	T49114	hypothetical prote
203	7	1.7	470	2	S33639	finger protein esc
204	7	1.7	479	2	T05588	cellulase (EC 3.2.
205	7	1.7	479	2	T07025	cellulase (EC 3.2.
206	7	1.7	481	2	T04021	cellulase (EC 3.2.
207	7	1.7	481	2	I49072	protein kinase - m
208	7	1.7	486	2	T06770	cellulase (EC 3.2.
209	7	1.7	486	2	T07885	cellulase (EC 3.2.
210	7	1.7	489	2	E86366	protein P26F24.6 [
211	7	1.7	489	2	T06350	cellulase (EC 3.2.
212	7	1.7	492	2	E86731	endo-1,4-beta-gluc
213	7	1.7	493	2	T06060	cellulase (EC 3.2.
214	7	1.7	494	2	T06059	cellulase (EC 3.2.
215	7	1.7	494	2	S11946	cellulase (EC 3.2.
216	7	1.7	501	2	A86158	endo-1,4-beta gluc
217	7	1.7	501	2	T52135	cellulase (EC 3.2.
218	7	1.7	502	2	A10151	probable bacteriop
219	7	1.7	505	2	T07883	cellulase (EC 3.2.
220	7	1.7	507	2	JC7226	endo-1,3(4)-beta-g
221	7	1.7	507	2	S46500	cellulase (EC 3.2.
222	7	1.7	508	2	T37224	hypothetical prote
223	7	1.7	510	2	T10734	cellulase (EC 3.2.
224	7	1.7	510	2	S57808	cellulase (EC 3.2.
225	7	1.7	515	2	T46610	cellulase (EC 3.2.
226	7	1.7	516	2	T01419	cellulase (EC 3.2.
227	7	1.7	519	1	S03564	aldehyde dehydroge
228	7	1.7	520	2	D70776	probable export pr
229	7	1.7	528	2	C85056	probable DNA-bindi
230	7	1.7	530	2	A70589	probable efpa prot
231	7	1.7	533	2	S18539	actVA-1 protein -
232	7	1.7	535	2	C87113	probable secreted
233	7	1.7	539	2	F70737	chaperonin groEL1
234	7	1.7	539	2	T50579	probable membrane
235	7	1.7	541	2	E86789	protein T23E18.10
236	7	1.7	542	2	S73749	transport system p
237	7	1.7	544	2	T34681	probable secreted
238	7	1.7	544	2	S58532	mask protein (trnx
239	7	1.7	551	2	R83015	hypothetical prote
240	7	1.7	551	2	B84213	hypothetical prote
241	7	1.7	558	1	S05532	gamma-glutamyltran
242	7	1.7	571	2	T01511	hypothetical prote
243	7	1.7	574	2	B83388	probable carbamoyl
244	7	1.7	582	2	B70389	translation elonga
245	7	1.7	584	2	D84264	hypothetical prote
246	7	1.7	593	2	T20650	hypothetical prote
247	7	1.7	597	1	G32039	chitinase (EC 3.2.
248	7	1.7	597	2	T25547	hypothetical prote

249	7	1.7	600	2	D83286	hypothetical prote
250	7	1.7	608	2	A64992	sensor protein Ato
251	7	1.7	610	2	T35797	secreted chitinase
252	7	1.7	616	2	B3586	C4-dicarboxylate t
253	7	1.7	621	2	D96032	hypothetical prote
254	7	1.7	622	2	T49426	Type 2C protein ph
255	7	1.7	624	2	B83386	hypothetical prote
256	7	1.7	627	2	B96527	protein P27015.28
257	7	1.7	631	2	T35234	probable secreted
258	7	1.7	631	2	T38167	electron transfer
259	7	1.7	640	2	A41726	homeotic protein B
260	7	1.7	642	2	S27806	homeotic protein B
261	7	1.7	645	2	H70783	hypothetical prote
262	7	1.7	649	2	T40646	probable protein i
263	7	1.7	659	2	E84176	DNA mismatch repai
264	7	1.7	677	2	T45682	hypothetical prote
265	7	1.7	681	2	S73550	DNA polymerase III
266	7	1.7	687	2	B81027	glycyl-tRNA synth
267	7	1.7	687	2	G81970	probable glycine-t
268	7	1.7	689	2	SYECGB	glycine-tRNA ligas
269	7	1.7	689	2	B91184	glycine tRNA synth
270	7	1.7	689	2	G86030	glycine-tRNA ligas
271	7	1.7	689	2	A80494	hypothetical prote
272	7	1.7	712	2	T27165	carbon starvation
273	7	1.7	721	2	S56580	probable carbon st
274	7	1.7	721	2	A98293	1-deoxy-D-xyulose
275	7	1.7	721	2	D86134	hypothetical prote
276	7	1.7	735	2	T08140	hypothetical ferredoxi
277	7	1.7	762	2	T14815	Na+/H+ antiporter
278	7	1.7	774	2	D83208	probable NADH dehy
279	7	1.7	789	2	A82688	phosphoenolpyruvat
280	7	1.7	789	2	H97469	ltp protein - huma
281	7	1.7	878	2	H83184	DNA polymerase B1
282	7	1.7	887	2	S57430	probable helicase
283	7	1.7	896	2	S57723	glycoprotein 130 -
284	7	1.7	901	2	B84210	interleukin-6 sign
285	7	1.7	906	2	G70767	hypothetical prote
286	7	1.7	917	2	I49699	villin 3 [imported
287	7	1.7	918	2	A44257	sarcosine oxidase
288	7	1.7	962	2	T22459	cellulase (EC 3.2.
289	7	1.7	966	2	T50668	middle wall protei
290	7	1.7	1005	2	B82969	S-layer protein pr
291	7	1.7	1062	2	H83966	adenylate cyclase
292	7	1.7	1070	2	S75712	delta endotoxin -
293	7	1.7	1084	2	A28555	urea amidolyase-re
294	7	1.7	1099	2	T14850	bunetanide-sensiti
295	7	1.7	1144	1	A39833	hypothetical prote
296	7	1.7	1186	2	T18210	probable SpoIIIF-f
297	7	1.7	1207	2	H87475	collagen alpha 1(I
298	7	1.7	1212	2	A57187	glutamate synthase
299	7	1.7	1230	2	T22458	microtubule-associ
300	7	1.7	1345	2	A87102	probable tape-meas
301	7	1.7	1466	1	C8H07L	extracellular matr
302	7	1.7	1530	2	B82085	hypothetical prote
303	7	1.7	1734	2	A54602	hypothetical prote
304	7	1.7	1787	2	AG3360	probable GFPase-ac
305	7	1.7	1788	2	AH1447	legumen protein 6
306	7	1.7	2035	2	T31110	probable hemagglu
307	7	1.7	2946	2	T00867	large repetitive p
308	7	1.7	2957	2	T33152	hypothetical prote
309	7	1.7	3079	1	RGBY12	cryptic fibrillar p
310	7	1.7	3436	2	S55659	antifreeze protein
311	7	1.7	3535	2	B83641	iron-sulfur protei
312	7	1.7	3624	2	AD0835	hypothetical prote
313	7	1.7	5627	2	C83339	hypothetical prote
314	7	1.7	5627	2	A60722	antifreeze protein
315	6	1.5	40	1	PDF18G	hypothetical prote
316	6	1.5	45	2	A05163	hypothetical prote
317	6	1.5	53	2	PC4101	hypothetical prote
318	6	1.5	55	2	C82796	hypothetical prote
319	6	1.5	57	2	A61381	hypothetical prote
320	6	1.5	59	2	A71845	hypothetical prote
321	6	1.5	60	2	S13476	ferredoxin 2[4Fe-4
322	6	1.5	60	2	A29667	pulmonary surfacta
323	6	1.5	61	2	A13331	cbim protein [impo
324	6	1.5	62	2	AD2476	hypothetical prote
325	6	1.5	63	2	S03864	hypothetical prote
326	6	1.5	65	2	H64324	ferredoxin 2[4Fe-4
327	6	1.5	65	2	A95330	hypothetical prote
328	6	1.5	66	2	C72153	PIL protein - vari
329	6	1.5	66	2	A36839	CIL protein - vari
330	6	1.5	68	2	T28459	hypothetical prote
331	6	1.5	68	2	UC7133	antioxidant protei
332	6	1.5	69	2	S60826	M protein precurs
333	6	1.5	71	2	H97137	hypothetical prote
334	6	1.5	72	2	A35731	Caz+-transporting
335	6	1.5	72	2	S68883	light-harvesting p
336	6	1.5	72	2	S74974	hypothetical prote
337	6	1.5	72	2	C71355	probable ribosomal
338	6	1.5	73	2	H81298	probable molybdopt
339	6	1.5	73	2	C82586	hypothetical prote
340	6	1.5	73	2	C95909	conserved hypotnet
341	6	1.5	77	2	D81267	50S ribosomal prot
342	6	1.5	78	2	S74715	ribosomal protein
343	6	1.5	78	2	AG2134	50S ribosomal prot
344	6	1.5	79	2	T36418	hypothetical prote
345	6	1.5	80	2	OSH07B	cytochrome-c oxida
346	6	1.5	80	2	E90686	hypothetical prote
347	6	1.5	80	2	A70688	hypothetical prote
348	6	1.5	80	2	A85537	unknown [imported]
349	6	1.5	81	2	H64814	molybdopterin bios
350	6	1.5	81	2	A82251	molybdopterin bios
351	6	1.5	81	2	E90736	molybdopterin bios
352	6	1.5	81	2	C05586	molybdopterin bios
353	6	1.5	83	2	AC0598	molybdopterin conv
354	6	1.5	83	2	D87676	hypothetical prote
355	6	1.5	83	2	T34904	hypothetical prote
356	6	1.5	83	2	T02136	hypothetical prote
357	6	1.5	84	2	T07115	metallothionein-11
358	6	1.5	84	2	D70672	hypothetical prote
359	6	1.5	85	2	S62832	cytochrome-c oxida
360	6	1.5	88	1	OSB07B	hypothetical prote
361	6	1.5	89	2	T45015	hypothetical prote
362	6	1.5	92	2	H90956	hypothetical prote
363	6	1.5	92	2	T36228	hypothetical prote
364	6	1.5	92	2	D85805	unknown protein en
365	6	1.5	93	2	C97819	hypothetical prote
366	6	1.5	95	2	C75478	conserved hypotnet
367	6	1.5	96	2	A86912	hypothetical prote
368	6	1.5	96	2	AG3001	50S ribosomal prot
369	6	1.5	98	1	AF3827	chaperonin groES -
370	6	1.5	98	2	AF3640	10K chaperonin gro
371	6	1.5	100	2	E86121	cytochrome b (562)
372	6	1.5	100	2	E91280	hypothetical prote
373	6	1.5	100	2	I40669	hypothetical prote
374	6	1.5	100	2	UQ0205	hypothetical 12.5K
375	6	1.5	101	2	H86885	50S ribosomal prot
376	6	1.5	102	2	C87282	conserved hypotnet
377	6	1.5	102	2	T35134	hypothetical prote
378	6	1.5	104	2	S75246	hypothetical prote
379	6	1.5	104	2	C82935	conserved hypotnet
380	6	1.5	105	2	F72662	hypothetical prote
381	6	1.5	105	2	G87591	hypothetical prote
382	6	1.5	106	2	S13027	cysteine proteinas
383	6	1.5	106	2	S37865	hypothetical prote
384	6	1.5	106	2	C83584	conserved hypotnet
385	6	1.5	107	1	PVRX2	pavabunibun beta l
386	6	1.5	107	2	A38375	oryzacystatin II -
387	6	1.5	107	2	B70650	probable emtr prot
388	6	1.5	107	2	AG3321	hypothetical prote
389	6	1.5	107	2	G81080	hypothetical prote
390	6	1.5	107	2	AF1965	hypothetical prote
391	6	1.5	108	2	I64217	ribosomal protein
392	6	1.5	108	2	S34941	fimbrial protein S
393	6	1.5	108	2	T40960	hypothetical prote
394	6	1.5	108	2	A48831	vitelline membrane

395	6	1.5	108	2	F72477	468	6	1.5	135	2	B49218	hemagglutinin homo
396	6	1.5	108	2	A95904	469	6	1.5	135	2	H72499	hypothetical prote
397	6	1.5	109	1	TNBOA1	470	6	1.5	135	2	G87686	hypothetical prote
398	6	1.5	109	2	T43705	471	6	1.5	135	2	JC4734	flagellar protein
399	6	1.5	109	2	G64609	472	6	1.5	135	2	T48927	hypothetical prote
400	6	1.5	110	1	PEBO	473	6	1.5	135	2	S67924	spore-wall fungal
401	6	1.5	110	2	B55863	474	6	1.5	136	2	C81226	hypothetical prote
402	6	1.5	110	2	T45387	475	6	1.5	136	2	B82787	hypothetical prote
403	6	1.5	111	2	S62830	476	6	1.5	137	2	S76236	hypothetical prote
404	6	1.5	112	2	D70535	477	6	1.5	137	2	E72614	hypothetical prote
405	6	1.5	113	2	S62528	478	6	1.5	138	1	BMD08	development-specif
406	6	1.5	113	2	S23439	479	6	1.5	138	2	T07203	hypothetical prote
407	6	1.5	113	2	JN0646	480	6	1.5	139	2	D70028	conserved hypotet
408	6	1.5	113	2	G64945	481	6	1.5	139	2	C70718	hypothetical prote
409	6	1.5	113	2	E90947	482	6	1.5	139	2	C87544	hypothetical prote
410	6	1.5	113	2	A85796	483	6	1.5	140	2	T49481	hypothetical prote
411	6	1.5	113	2	AH0397	484	6	1.5	140	2	AD2423	hypothetical prote
412	6	1.5	113	2	T45276	485	6	1.5	141	2	C69007	ribosomal protein
413	6	1.5	114	2	G81353	486	6	1.5	141	2	T34890	probable IS1648 tr
414	6	1.5	115	2	B75535	487	6	1.5	142	2	C32223	interleukin-7 prec
415	6	1.5	116	2	T37186	488	6	1.5	142	2	G72050	conserved hypotet
416	6	1.5	116	2	B75456	489	6	1.5	142	2	D86574	CTS50 hypothetical
417	6	1.5	116	2	S52219	490	6	1.5	143	2	AC2693	transcription regu
418	6	1.5	116	2	D84285	491	6	1.5	144	2	DB3152	hypothetical prote
419	6	1.5	117	2	F95862	492	6	1.5	144	2	C96004	conserved hypotet
420	6	1.5	118	2	F87447	493	6	1.5	144	2	*B70949	hypothetical prote
421	6	1.5	119	1	CUPSAM	494	6	1.5	144	2	G64423	hypothetical prote
422	6	1.5	119	2	SL7123	495	6	1.5	145	2	F69214	heat shock protein
423	6	1.5	119	2	C90864	496	6	1.5	145	2	C84610	probable beta-hydr
424	6	1.5	119	2	F85754	497	6	1.5	145	2	F83267	conserved hypotet
425	6	1.5	119	2	C82795	498	6	1.5	146	1	HBAK	hemoglobin beta ch
426	6	1.5	119	2	H97183	499	6	1.5	146	1	HBAQ	hemoglobin beta ch
427	6	1.5	120	1	H64900	500	6	1.5	146	2	G7474	probable asc-fami
428	6	1.5	120	2	T35920	501	6	1.5	146	2	T10511	hypothetical prote
429	6	1.5	121	2	T03318	502	6	1.5	146	2	AE3630	transcription regu
430	6	1.5	121	2	S34233	503	6	1.5	147	2	C31844	K118 protein - Str
431	6	1.5	121	2	F83940	504	6	1.5	148	2	D75276	response regulator
432	6	1.5	122	2	D42645	505	6	1.5	148	2	AC1122	hypothetical prote
433	6	1.5	122	2	B86570	506	6	1.5	149	2	D53399	protein (imported
434	6	1.5	122	2	E81664	507	6	1.5	150	2	G87590	cytochrome c-type
435	6	1.5	122	2	G72054	508	6	1.5	150	2	A70305	(3R)-hydroxymyrist
436	6	1.5	123	2	G97885	509	6	1.5	150	2	S58172	mitramycin polyke
437	6	1.5	124	2	B70148	510	6	1.5	150	2	D87652	hypothetical prote
438	6	1.5	124	2	F71347	511	6	1.5	151	2	A24475	superoxide dismuta
439	6	1.5	124	2	T05467	512	6	1.5	151	2	AC2084	phosphonate metabo
440	6	1.5	125	2	T50866	513	6	1.5	151	2	T15428	hypothetical prote
441	6	1.5	125	2	T34695	514	6	1.5	152	2	A83053	conserved hypotet
442	6	1.5	125	2	AF1932	515	6	1.5	152	2	A87679	conserved hypotet
443	6	1.5	126	2	T21448	516	6	1.5	153	2	S21767	hypothetical prote
444	6	1.5	126	2	H70634	517	6	1.5	153	2	F69267	cytochrome-c oxida
445	6	1.5	126	2	PN0542	518	6	1.5	154	2	B83304	hypothetical prote
446	6	1.5	127	2	B64034	519	6	1.5	154	2	AC3461	hypothetical prote
447	6	1.5	127	2	B75353	520	6	1.5	154	2	D97415	hypothetical prote
448	6	1.5	127	2	A72712	521	6	1.5	155	2	C64921	outer membrane lip
449	6	1.5	128	1	CBEBC62	522	6	1.5	155	2	F90922	outer membrane mem
450	6	1.5	128	2	AF1056	523	6	1.5	155	2	B85771	probable outer mem
451	6	1.5	128	2	B70647	524	6	1.5	155	2	A10693	outer membrane lip
452	6	1.5	128	2	T50723	525	6	1.5	155	2	H86972	riboflavin synthas
453	6	1.5	128	2	S72600	526	6	1.5	155	2	D87399	cytochrome c fam1
454	6	1.5	129	2	T16369	527	6	1.5	156	2	A97397	deoxyuridine 5'-tr
455	6	1.5	130	2	T34729	528	6	1.5	156	2	AB2615	deoxyuridine 5'-tri
456	6	1.5	131	2	B72523	529	6	1.5	156	2	AH1606	acetyl-CoA carboxy
457	6	1.5	131	2	AB1525	530	6	1.5	156	2	C97503	hypothetical prote
458	6	1.5	131	2	A87355	531	6	1.5	156	2	*S08427	12k protein ma12A
459	6	1.5	131	2	S18540	532	6	1.5	156	2	AB0079	hypothetical prote
460	6	1.5	132	2	G82423	533	6	1.5	157	2	F91172	probable phosphor
461	6	1.5	133	2	H87681	534	6	1.5	157	2	F66018	probable phosphor
462	6	1.5	134	1	F71142	535	6	1.5	157	2	AB2539	hypothetical prote
463	6	1.5	134	2	AC1166	536	6	1.5	158	2	T06078	hypothetical prote
464	6	1.5	134	2	AC1943	537	6	1.5	158	2	B91251	hypothetical prote
465	6	1.5	134	2	D97582	538	6	1.5	158	2	H85620	hypothetical prote
466	6	1.5	134	2	AD2803	539	6	1.5	159	2	H75391	hypothetical prote
467	6	1.5	135	2	D84730	540	6	1.5	160	2	F75450	hypothetical prote

687	6	1.5	201	2	T06679	6	1.5	213	2	T17334	6	1.5	213	2	T17334	hypothetical prote
688	6	1.5	201	2	S55000	6	1.5	213	2	AF2664	6	1.5	213	2	AF2664	ATP synthase B' ch
689	6	1.5	201	2	G02312	6	1.5	214	2	D4681	6	1.5	214	2	D4681	glutathione transf
690	6	1.5	201	2	G70538	6	1.5	214	2	F70784	6	1.5	214	2	F70784	hypothetical prote
691	6	1.5	201	2	G71428	6	1.5	214	2	T20056	6	1.5	214	2	T20056	hypothetical prote
692	6	1.5	201	2	B72739	6	1.5	215	2	C46681	6	1.5	215	2	C46681	glutathione transf
693	6	1.5	202	2	H82191	6	1.5	215	2	T36281	6	1.5	215	2	T36281	probable hydrolase
694	6	1.5	202	2	T23654	6	1.5	215	2	T44346	6	1.5	215	2	T44346	hypothetical prote
695	6	1.5	202	2	T23654	6	1.5	215	2	F71368	6	1.5	215	2	F71368	hypothetical prote
696	6	1.5	202	2	AC2714	6	1.5	215	2	D84382	6	1.5	215	2	D84382	hypothetical prote
697	6	1.5	202	2	H97495	6	1.5	215	2	E97446	6	1.5	215	2	E97446	ATP synthase chain
698	6	1.5	202	2	A87636	6	1.5	215	2	G90164	6	1.5	215	2	G90164	uracil phosphorib
699	6	1.5	202	2	U00138	6	1.5	216	2	G92163	6	1.5	216	2	G92163	5-methylthiodenos
700	6	1.5	203	2	G70752	6	1.5	217	2	F71293	6	1.5	217	2	F71293	hypothetical prote
701	6	1.5	203	2	I40111	6	1.5	217	2	F71953	6	1.5	217	2	F71953	hypothetical prote
702	6	1.5	204	2	H70522	6	1.5	217	2	C96748	6	1.5	217	2	C96748	unknown protein T1
703	6	1.5	204	2	B31227	6	1.5	218	2	T34650	6	1.5	218	2	T34650	probable transfera
704	6	1.5	204	2	E95345	6	1.5	218	2	AB0254	6	1.5	218	2	AB0254	probable fumarylac
705	6	1.5	204	2	I40112	6	1.5	218	2	BE1111	6	1.5	218	2	BE1111	nickel-dependent h
706	6	1.5	205	2	I50480	6	1.5	219	2	B81906	6	1.5	219	2	B81906	probable membrane
707	6	1.5	205	2	B87296	6	1.5	219	2	B29224	6	1.5	219	2	B29224	GTP-binding protei
708	6	1.5	205	2	I40105	6	1.5	219	2	T50020	6	1.5	219	2	T50020	(3R)-hydroxymyrist
709	6	1.5	205	2	E84334	6	1.5	219	2	AE0724	6	1.5	219	2	AE0724	probable hydrolase
710	6	1.5	205	2	C83521	6	1.5	220	2	AE2535	6	1.5	220	2	AE2535	transcription regu
711	6	1.5	205	2	AB3600	6	1.5	220	2	F70906	6	1.5	220	2	F70906	probable O-methyl
712	6	1.5	206	2	B97594	6	1.5	221	2	F64327	6	1.5	221	2	F64327	H+-transporting tw
713	6	1.5	206	2	AI2123	6	1.5	221	2	AG1595	6	1.5	221	2	AG1595	weakly oligopeptid
714	6	1.5	206	2	T27766	6	1.5	222	2	AG0164	6	1.5	222	2	AG0164	arginine transport
715	6	1.5	206	2	AE2222	6	1.5	222	2	AE0607	6	1.5	222	2	AE0607	butyrate-acetoacet
716	6	1.5	207	2	S31011	6	1.5	223	2	G64158	6	1.5	223	2	G64158	hypothetical prote
717	6	1.5	207	2	E97825	6	1.5	223	2	D70526	6	1.5	223	2	D70526	secretory protein
718	6	1.5	207	2	I40271	6	1.5	224	2	B40850	6	1.5	224	2	B40850	phosphoribosylform
719	6	1.5	207	2	I40276	6	1.5	224	2	BD4345	6	1.5	224	2	BD4345	serum amyloid P-co
720	6	1.5	207	2	S69919	6	1.5	224	2	A30528	6	1.5	224	2	A30528	probable urease ac
721	6	1.5	207	2	S69924	6	1.5	224	2	D70665	6	1.5	224	2	D70665	ABC transporter, p
722	6	1.5	207	2	S37727	6	1.5	224	2	AB1377	6	1.5	224	2	AB1377	ABC transporter, p
723	6	1.5	207	2	B75294	6	1.5	224	2	AD1746	6	1.5	224	2	AD1746	hypothetical prote
724	6	1.5	207	2	B70708	6	1.5	224	2	A95394	6	1.5	224	2	A95394	TIN6.14 protein -
725	6	1.5	208	2	T48860	6	1.5	225	2	G86148	6	1.5	225	2	G86148	conserved hypothet
726	6	1.5	208	2	F75427	6	1.5	225	2	AE0610	6	1.5	225	2	AE0610	urase accessory p
727	6	1.5	208	2	T36390	6	1.5	225	2	T50707	6	1.5	225	2	T50707	probable DNA-3-met
728	6	1.5	208	2	AB3406	6	1.5	225	2	A75257	6	1.5	225	2	A75257	beta-phosphoglucom
729	6	1.5	209	2	XUFP11	6	1.5	226	2	E70034	6	1.5	226	2	E70034	conserved hypothet
730	6	1.5	209	2	I40281	6	1.5	226	2	AE0967	6	1.5	226	2	AE0967	hypothetical prote
731	6	1.5	209	2	S72929	6	1.5	226	2	T14675	6	1.5	226	2	T14675	probable transmemb
732	6	1.5	209	2	B95116	6	1.5	227	2	E70501	6	1.5	227	2	E70501	DNA gyrase chain A
733	6	1.5	209	2	H97985	6	1.5	227	2	S35735	6	1.5	227	2	S35735	photoresystem I chai
734	6	1.5	210	2	E69027	6	1.5	227	2	S94133	6	1.5	227	2	S94133	probable membrane
735	6	1.5	210	2	F71308	6	1.5	227	2	H86940	6	1.5	227	2	H86940	conserved hypothet
736	6	1.5	210	2	AC0667	6	1.5	227	2	G75269	6	1.5	227	2	G75269	heat shock protein
737	6	1.5	210	2	G95858	6	1.5	227	2	S56941	6	1.5	227	2	S56941	hypothetical prote
738	6	1.5	210	2	D89825	6	1.5	228	2	T51147	6	1.5	228	2	T51147	hypothetical prote
739	6	1.5	210	2	D84238	6	1.5	228	2	C70623	6	1.5	228	2	C70623	probable IS1560 tr
740	6	1.5	210	2	S69920	6	1.5	228	2	F70624	6	1.5	228	2	F70624	hypothetical prote
741	6	1.5	210	2	S69925	6	1.5	228	2	T24592	6	1.5	228	2	T24592	hypothetical prote
742	6	1.5	210	2	F84211	6	1.5	229	2	S72165	6	1.5	229	2	S72165	gene 7 protein - p
743	6	1.5	210	2	C69652	6	1.5	229	2	C43330	6	1.5	229	2	C43330	hypothetical prote
744	6	1.5	210	2	I58391	6	1.5	229	2	C68809	6	1.5	229	2	C68809	hypothetical prote
745	6	1.5	211	2	S69929	6	1.5	229	2	F75482	6	1.5	229	2	F75482	transcription regu
746	6	1.5	211	2	D69413	6	1.5	230	2	UVFUS	6	1.5	230	2	UVFUS	cutinase (EC 3.1.-
747	6	1.5	212	2	E82061	6	1.5	230	2	T36726	6	1.5	230	2	T36726	ubiquinol-cytochro
748	6	1.5	212	2	T05721	6	1.5	231	2	RDNCUP	6	1.5	231	2	RDNCUP	recombination prot
749	6	1.5	212	2	T10230	6	1.5	231	2	PC1108	6	1.5	231	2	PC1108	two components res
750	6	1.5	212	2	S44805	6	1.5	231	2	AD1785	6	1.5	231	2	AD1785	transcription regu
751	6	1.5	212	2	S42396	6	1.5	231	2	AF3037	6	1.5	231	2	AF3037	probable regulator
752	6	1.5	212	2	H87018	6	1.5	232	2	T36726	6	1.5	232	2	T36726	hypothetical prote
753	6	1.5	212	2	S23934	6	1.5	232	2	T14939	6	1.5	232	2	T14939	hypothetical prote
754	6	1.5	212	2	AC3048	6	1.5	233	2	D70537	6	1.5	233	2	D70537	conserved hypothet
755	6	1.5	212	2	H98237	6	1.5	233	2	BE6913	6	1.5	233	2	BE6913	glyceraldehyde-3-P
756	6	1.5	213	2	S26035	6	1.5	233	2	B94796	6	1.5	233	2	B94796	hypothetical prote
757	6	1.5	213	2	T02871	6	1.5	233	2	C75290	6	1.5	233	2	C75290	hypothetical prote
758	6	1.5	213	2	E75855	6	1.5	233	2	S57924	6	1.5	233	2	S57924	macrophage infecti
759	6	1.5	213	2	B70707	6	1.5	234	1	F65080	6	1.5	234	1	F65080	hypothetical prote

833	6	1.5	234	2	C91107	hypothetical prote	906	6	1.5	249	2	C84185	hypothetical prote
834	6	1.5	234	2	F85952	hypothetical prote	907	6	1.5	249	2	B87309	hypothetical prote
835	6	1.5	234	2	AD0878	conserved hypochet	908	6	1.5	249	2	B87505	ABC transporter, A
836	6	1.5	234	2	B83065	conserved hypochet,	909	6	1.5	249	2	B87597	transcription regu
837	6	1.5	234	2	154074	upstream stimulat	910	6	1.5	249	2	F87375	hypothetical prote
838	6	1.5	234	2	C87713	conserved hypochet	911	6	1.5	249	2	E98248	hypothetical prote
839	6	1.5	234	2	T35448	hypothetical prote	912	6	1.5	250	2	AP1636	N-acetylglutamate
840	6	1.5	234	2	B69037	conserved hypochet	913	6	1.5	250	2	H90400	hypothetical prote
841	6	1.5	235	2	H70556	hypothetical prote	914	6	1.5	250	2	H86637	conserved hypochet
842	6	1.5	235	2	S39652	secretion protein	915	6	1.5	250	2	JC7339	maleate isomerase
843	6	1.5	235	2	A99186	hypothetical prote	916	6	1.5	251	2	S28171	H ⁺ -transporting tw
844	6	1.5	235	2	A13100	phosphoribosylran	917	6	1.5	251	2	E90206	hypothetical prote
845	6	1.5	235	2	AB2613	orotidine 5'-monop	918	6	1.5	251	2	D90243	conserved hypochet
846	6	1.5	236	1	T46967	diheme cytochrome	919	6	1.5	251	2	T03535	precorrin-6x reduc
847	6	1.5	236	2	D71376	hypothetical prote	920	6	1.5	252	2	D82667	transcription regu
848	6	1.5	236	2	H95939	conserved hypochet	921	6	1.5	252	2	C81877	transcription regu
849	6	1.5	236	2	AH3247	uracil-DNA glycosy	922	6	1.5	252	2	S50806	hypothetical prote
850	6	1.5	237	2	AE2878	anti-sigma cross-r	923	6	1.5	253	2	F83134	probable short-cha
851	6	1.5	237	2	UC4914	spiralin - Splropl	924	6	1.5	253	2	C72758	transcription regu
852	6	1.5	237	2	A36149	6-phosphogluconol	925	6	1.5	253	2	F87371	transcription fact
853	6	1.5	237	2	F95271	hypothetical prote	926	6	1.5	253	2	A38201	hypothetical prote
854	6	1.5	237	2	AG3573	hypothetical prote	927	6	1.5	253	2	AG2627	hypothetical prote
855	6	1.5	238	2	AE2373	hypothetical prote	928	6	1.5	253	2	AB2639	hypothetical prote
856	6	1.5	238	2	A72577	hypothetical prote	929	6	1.5	253	2	H81286	probable sugar nuc
857	6	1.5	238	2	T49297	hypothetical prote	930	6	1.5	254	1	BV8122	PF112 protein pre
858	6	1.5	238	2	D82619	glucose inhibited	931	6	1.5	254	2	D90929	probable transpor
859	6	1.5	239	2	A44343	promotes sex organ	932	6	1.5	254	2	S34724	probable oxidoredu
860	6	1.5	239	2	T51534	hypothetical prote	933	6	1.5	254	2	AB3345	probable thioester
861	6	1.5	239	2	A75585	transcription regu	934	6	1.5	254	2	A97263	lactose phosphotra
862	6	1.5	239	2	T36572	hypothetical prote	935	6	1.5	254	2	T36161	hypothetical prote
863	6	1.5	240	2	E72170	A56r protein - var	936	6	1.5	254	2	H85777	probable transpor
864	6	1.5	240	2	S61872	hypothetical prote	937	6	1.5	255	2	T23488	hypothetical prote
865	6	1.5	241	1	QOCVW2	coat protein - abu	938	6	1.5	255	2	T03519	coalbumin (5'-phos
866	6	1.5	241	2	G97654	uracil-DNA glycosy	939	6	1.5	255	2	B82729	undecaprenyl pyro
867	6	1.5	241	2	D87494	conserved hypochet	940	6	1.5	255	2	T50916	hypothetical prote
868	6	1.5	241	2	A70874	probable exported	941	6	1.5	255	2	H64648	alternative transc
869	6	1.5	241	2	F95355	protein [imported	942	6	1.5	255	2	F71938	RNA polymerase sig
870	6	1.5	241	2	AH3543	gtp-binding protei	943	6	1.5	255	2	A40613	avirulence protein
871	6	1.5	242	1	MFIVCJ	matrix protein M1	944	6	1.5	255	2	C97081	2-Hydroxy-6-Oxo-6-
872	6	1.5	242	2	AC1542	B. subtilis NagB p	945	6	1.5	255	2	H97394	orotidine 5'-monop
873	6	1.5	242	2	D87346	transcription regu	946	6	1.5	255	2	AD1946	hypothetical prote
874	6	1.5	242	2	C83230	hypothetical prote	947	6	1.5	256	2	T31268	4-oxaloacetonate d
875	6	1.5	242	2	F84315	cobalt transport A	948	6	1.5	256	2	T35795	hypothetical prote
876	6	1.5	242	2	E70248	hypothetical prote	949	6	1.5	256	2	D83336	probable thiol-dis
877	6	1.5	242	2	A81782	outer membrane pro	950	6	1.5	256	2	G70666	hypothetical prote
878	6	1.5	242	2	C81205	outer membrane pro	951	6	1.5	256	2	F47031	dialphos biosynth
879	6	1.5	243	2	E70846	hypothetical prote	952	6	1.5	256	2	T48588	hypothetical prote
880	6	1.5	243	2	B58585	probable transcrip	953	6	1.5	257	2	AD2934	short chain dehydr
881	6	1.5	243	2	H86487	hypothetical prote	954	6	1.5	257	2	AG2750	3-hydroxyacyl-CoA
882	6	1.5	243	2	AH2640	transcription regu	955	6	1.5	257	2	C98348	reductase (Alu13282
883	6	1.5	243	2	B97423	transcription acti	956	6	1.5	257	2	F97531	probable short-cha
884	6	1.5	244	2	G95970	probable transcrip	957	6	1.5	257	2	AH3584	cystine-binding pe
885	6	1.5	244	2	S48682	hypothetical prote	958	6	1.5	257	2	B86387	conserved hypochet
886	6	1.5	245	2	AF1070	probable DNA repli	959	6	1.5	257	2	S16865	gene F41 protein -
887	6	1.5	245	2	H72612	hypothetical prote	960	6	1.5	257	2	B84346	hypothetical prote
888	6	1.5	245	2	E86536	hypothetical prote	961	6	1.5	257	2	B75510	hypothetical prote
889	6	1.5	245	2	G72086	hypothetical prote	962	6	1.5	257	2	C89845	conserved hypochet
890	6	1.5	246	1	S30274	proteasome endopep	963	6	1.5	258	2	G84858	floral homeodomain
891	6	1.5	246	1	UX0230	proteasome endopep	964	6	1.5	258	2	T52377	PR1B5 protein [imp
892	6	1.5	246	2	H95408	probable short cha	965	6	1.5	258	2	D95910	probable membrane-
893	6	1.5	246	2	AB2062	diaminopimelate ep	966	6	1.5	258	2	JC2667	UDP-N-acetylmuramo
894	6	1.5	246	2	B69766	transcription regu	967	6	1.5	258	2	T25958	hypothetical prote
895	6	1.5	246	2	E39534	floral homeotic pr	968	6	1.5	258	2	AI0593	ABC transporter AT
896	6	1.5	247	1	A25852	trypsin (EC 3.4.21	969	6	1.5	259	2	T34637	probable type IV p
897	6	1.5	247	2	AD0757	cobalamn (5'-phos	970	6	1.5	259	2	B69113	cell division inh
898	6	1.5	247	2	S60307	fbp6 protein - gar	971	6	1.5	259	2	AI13501	orotidine 5'-phosp
899	6	1.5	247	2	T52511	related to lysopho	972	6	1.5	260	2	G87408	gluconate 5-dehydr
900	6	1.5	247	2	G90348	hypothetical prote	973	6	1.5	260	2	D96004	probable 3-oxoacyl
901	6	1.5	248	1	PAORCS	acid phosphatase (974	6	1.5	260	2	AI0934	glutamate racemase
902	6	1.5	248	2	B87425	transcription regu	975	6	1.5	260	2	B83675	enoyl-CoA hydrat
903	6	1.5	248	2	A39534	floral homeotic pr	976	6	1.5	260	2	B87392	conserved hypochet
904	6	1.5	249	1	S19187	acid phosphatase (977	6	1.5	261	2	C81368	probable glutamine
905	6	1.5	249	2	AB2262	hypothetical prote	978	6	1.5	261	2	S72895	hypothetical prote

979 6 1.5 262 2 AC2644 flagellar basal-bo
980 6 1.5 262 2 C97426 flagellar basal-bo
981 6 1.5 262 2 B82950 chromosome partiti
982 6 1.5 262 2 A75009 probable aryl phos
983 6 1.5 262 2 C70587 hypothetical prote
984 6 1.5 262 2 T19323 hypothetical prote
985 6 1.5 262 2 B70699 hypothetical prote
986 6 1.5 263 1 SNRCC2 proteasome endopep
987 6 1.5 263 2 A83956 flagellar hook pro
988 6 1.5 263 2 H87296 aminomethyltransfe
989 6 1.5 263 2 C84598 probable ER lumen
990 6 1.5 264 2 B69535 DNA-directed RNA p
991 6 1.5 264 2 H83170 conserved hypothet
992 6 1.5 264 2 B70546 probable pnp prote
993 6 1.5 264 2 T34953 sporulation protei
994 6 1.5 264 2 G87356 hypothetical prote
995 6 1.5 264 2 AD3298 sensory transducti
996 6 1.5 265 1 JC5530 T-cluster binding
997 6 1.5 265 1 G69084 conserved hypothet
998 6 1.5 265 2 C82358 glutamate racemase
999 6 1.5 265 2 T07800 inositol-1(or 4)-m
1000 6 1.5 265 2 C84354 hypothetical prote

ALIGNMENTS

RESULT 1

membrane fusion protein NMA1970 [imported] - Neisseria meningitidis (strain Z2491 serogr
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: H81825
R:Parkhill, J., Achtman, M., James, K.D., Bentley, S.D., Churcher, C., Klee, S.R., Morel
R., Holtroyd, S., Jorgensen, K., Leather, S., Mouton, S., Mungall, K., Quail, M.A., Rajandream,
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A:Reference number: A81775; MUID:20222556; PMID:10761919
A:Accession: H81825
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-412 <PAR>
A:Cross-references: GB:AL162757; GB:AL157959; NID:G7380371; PIDN:CAB85190.1; PID:G738060
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: mtrC; NMA1970

Query Match 41.5%; Score 171; DB 2; Length 412;
Best Local Similarity 99.6%; Pred. No. 2,6e-159;
Matches 271; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAFYAFKMRRAAALAAVALVYSSCGKGDAAOGGPPAREAPAPVGVVTHPQTVALT 60
DB 1 MAFYAFKMRRAAALAAVALVYSSCGKGDAAOGGPPAREAPAPVGVVTHPQTVALT 60
QY 61 VELPGLESLRTADYVAOVGGIIOKRLFOEGSVVRAGOPLYOIDSSTVEANTESARAQLA 120
DB 61 VELPGLESLRTADYVAOVGGIIOKRLFOEGSVVRAGOPLYOIDSSTVEANTESARAQLA 120
QY 121 TAQATLAKADADLARYKPLVAEAVSRQYDAAVTAKRSAGGVKAAQAISAGINLNR 180
DB 121 TAQATLAKADADLARYKPLVAEAVSRQYDAAVTAKRSAGGVKAAQAISAGINLNR 180
QY 181 SRTITPISGFIQSQKSEGTLLNAGDTYLATIRQNPVYVWVQSASVWMLRQIAG 240
DB 181 SRTITPISGFIQSQKSEGTLLNAGDTYLATIRQNPVYVWVQSASVWMLRQIAG 240
QY 241 KLLADGVAVANGIKPDDGTVYPEKGRLLPADP 272
DB 241 KLLADGVAVANGIKPDDGTVYPEKGRLLPADP 272

RESULT 2

F81051
membrane fusion protein NMB1716 [imported] - Neisseria meningitidis (strain MCS8 serogr
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: F81051
R:Teitelin, H., Saunders, N.J., Heideberg, J., Jeffries, A.C., Nelson, K.E., Eisen, J.A
Hickey, E.K., Haft, D.H., Salzberg, S.L., White, O., Fleischmann, R.D., Dougherty, B.A.,
Li, H., Qin, H., Vamathevan, J., Gill, J., Scarlato, V., Maignani, V., Piazza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G., Sun, L., Smith, H.O., Fraser, C.M., Moxon, E.R., Rappuoli, R., Ve
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MCS8.
A:Reference number: A81000; MUID:20175753; PMID:10710307
A:Accession: F81051
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-412 <TER>
A:Cross-references: GB:AE002521; GB:AE002098; NID:G7226962; PIDN:AAF42063.1; PID:G722697
A:Experimental source: serogroup B, strain MCS8
C:Genetics:
A:Gene: NMB1716

Query Match 32.3%; Score 133; DB 2; Length 412;
Best Local Similarity 100.0%; Pred. No. 4.8e-122;
Matches 133; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 274 VNESTGQITLRAAVPNDQIIMPGLYVRLMDQVAVDPAFVPPQAVTRGAKDTVMIVNA 333
DB 274 VNESTGQITLRAAVPNDQIIMPGLYVRLMDQVAVDPAFVPPQAVTRGAKDTVMIVNA 333
QY 334 OGMEPREVTVAAOOGGTWYITSGIKDQDKVYVBSISAGITGAKKTPKEMASSENQAA 393
DB 334 OGMEPREVTVAAOOGGTWYITSGIKDQDKVYVBSISAGITGAKKTPKEMASSENQAA 393
QY 394 APOGSGVTASEAK 406
DB 394 APOGSGVTASEAK 406

RESULT 3

S42418
Probable antibiotic resistance protein mtrC - Neisseria gonorrhoeae (strain CH95) (fragm
C:Species: Neisseria gonorrhoeae
C:Date: 07-May-1998 #sequence_revision 15-May-1998 #text_change 08-Oct-1999
C:Accession: S42418; S40252
R:Pan, W., Spratt, B.G.
Mol. Microbiol. 11, 769-775, 1994
A:Title: Regulation of the permeability of the gonococcal cell envelope by the mtr syste
A:Reference number: S42417; MUID:94254732; PMID:8196548
A:Accession: S42418
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA

A:Residues: 1-271 <PAR>
A:Cross-references: EMBL:Z25796; NID:G438190; PIDN:CAA81046.1; PID:G438192
A:Experimental source: CH95
C:Genetics:
A:Gene: mtrC

C:Superfamily: lipoyl/biotin-binding homology
F:68-111,183-211/Domain: lipoyl/biotin-binding homology #status atypical <LPB>

Query Match 24.3%; Score 100; DB 2; Length 271;
Best Local Similarity 100.0%; Pred. No. 7.7e-90;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 OEYDAAVTAKSAGVGAQAQAISAGINRSTRITPISGFIQSQKSEGTLLNAGDT 207
DB 148 OEYDAAVTAKSAGVGAQAQAISAGINRSTRITPISGFIQSQKSEGTLLNAGDT 207
QY 208 TVLATIRQNPVYVWVQSASVWMLRQIAGKLLAADG 247
DB 208 TVLATIRQNPVYVWVQSASVWMLRQIAGKLLAADG 247

RESULT 4

E85757
 Probable efflux pump Z2509 [imported] - Escherichia coli (strain O157:H7, substrain EDL9
 C/Species: Escherichia coli
 C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C/Accession: E85757
 R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaesner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoudis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A/Reference number: A85480; MUID:21074935; PMID:11206551
 A/Accession: E85757
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-373 <STO>
 A/Cross-references: GB:AE005174; NID:g1515494; PIDN:AA656521.1; GSPDB:GN00145; UWGP:Z25
 A/Experimental source: strain O157:H7, substrain EDL933
 C/Genetics:
 A/Gene: Z2509

Query Match 2.7%; Score 11; DB 2; Length 373;
 Best Local Similarity 100.0%; Pred. No. 0.021;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 78 QVGGIIQKRLF 88
 |||||
 Db 63 QVGGIIQKRLF 73

RESULT 5

G90861
 Probable efflux pump Ecs1863 [similarity] - Escherichia coli (strain O157:H7, substrain
 C/Species: Escherichia coli
 C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 02-Nov-2001
 C/Accession: G90861
 R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
 Sasawara, N.; Yasunaga, T.; Kuwara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
 A/Reference number: A99629; MUID:21156231; PMID:11258796
 A/Accession: G90861
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-373 <HAY>
 A/Cross-references: GB:BA000007; PIDN:BA835286.1; PID:g13361328; GSPDB:GN00154
 A/Experimental source: strain O157:H7, substrain RMD 0509952
 C/Genetics:
 A/Gene: Ecs1863

Query Match 2.7%; Score 11; DB 2; Length 373;
 Best Local Similarity 100.0%; Pred. No. 0.021;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 78 QVGGIIQKRLF 88
 |||||
 Db 63 QVGGIIQKRLF 73

RESULT 6

E87349
 HlyD family secretion protein [imported] - Caulobacter crescentus
 C/Species: Caulobacter crescentus
 C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
 C/Accession: E87349
 R/Nierman, W.C.; Feldlym, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
 B.; Land, M.T.; DeBoy, R.T.; Dodson, R.U.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolor
 n, J.; Brmlaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A/Title: Complete Genome Sequence of Caulobacter crescentus.
 A/Reference number: A87249; MUID:21173698; PMID:11259647
 A/Accession: E87349
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-388 <STO>

A/Cross-references: GB:AE005673; NID:g13422055; PIDN:AAK22793.1; GSPDB:GN00148
 C/Genetics:
 A/Gene: CC0808

Query Match 2.7%; Score 11; DB 2; Length 388;
 Best Local Similarity 100.0%; Pred. No. 0.021;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 94 VRAGQPLXQID 104
 |||||
 Db 89 VRAGQPLXQID 99

RESULT 7

AH0346
 putative HlyD family secretion protein [imported] - Yersinia pestis (strain CO92)
 C/Species: Yersinia pestis
 C/Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
 C/Accession: AH0346
 R/Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tibball, R.W.; Holden, M.T.G.; Prentice, M.B.
 deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
 il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
 Nature 413, 523-527, 2001
 A/Title: Genome sequence of Yersinia pestis, the causative agent of plague.
 A/Reference number: AB0001; MUID:21470413; PMID:11586360.
 A/Accession: AH0346
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-444 <KTR>
 A/Cross-references: GB:AL590842; PIDN:CAC92099.1; PID:g15980817; GSPDB:GN00175
 C/Genetics:
 A/Gene: YPO2847

Query Match 2.2%; Score 9; DB 2; Length 444;
 Best Local Similarity 100.0%; Pred. No. 2.2;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 181 SRTAPISG 189
 |||||
 Db 198 SRTAPISG 206

RESULT 8

C81232
 50S ribosomal protein L24 NMB0153 [imported] - Neisseria meningitidis (strain MC58 serog
 C/Species: Neisseria meningitidis
 C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Feb-2001
 C/Accession: C81232; D82004
 R/Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A
 Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
 ri, H.; Qin, H.; Vamathevan, J.; Gill, U.; Scarlato, V.; Maignani, V.; Pizzi, M.
 Science 287, 1809-1815, 2000
 A/Authors: Grandl, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
 A/Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
 A/Reference number: A81000; MUID:20175755; PMID:10710307
 A/Accession: C81232
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-107 <TEP>
 A/Cross-references: GB:AE002373; GB:AE002098; NID:g7225359; PIDN:AAE40611.1; PID:g722537
 A/Experimental source: serogroup B, strain MC58
 R/Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churchev, C.; Klee, S.R.; Morel
 Holroyd, S.; Jørgen, K.; Leach, S.; Moutle, S.; Mungall, K.; Quail, M.A.; Rajandream,
 Nature 404, 502-506, 2000
 A/Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491.
 A/Reference number: A81775; MUID:20222556; PMID:10761919
 A/Accession: D82004
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-107 <PAR>
 A/Cross-references: GB:AL162752; GB:AL157959; NID:g7378778; PIDN:CAE83433.1; PID:g737889
 A/Experimental source: serogroup A, strain 22491
 C/Genetics:

A:Gene: rplX; NMB0153; NMA0118
C:Superfamily: Escherichia coli ribosomal protein L24

Query Match 1.9%; Score 8; DB 2; Length 107;
Best Local Similarity 100.0%; Pred. No. 5.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 DKRVVEGI 368
DB 29 GDKVVEVG 36

RESULT 9

ribosomal protein L24 - Chlamydia trachomatis

C:Species: Chlamydia trachomatis
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 13-Aug-1999

C:Accession: E42645; C71506

R:Kaul, R.; Gray, G.J.; Koehncke, N.R.; Gu, L.J.
J. Bacteriol. 174, 1205-1212, 1992

A:Title: Cloning and sequence analysis of the Chlamydia trachomatis spc ribosomal protei
A:Reference number: A42645; MUID:92138612; PMID:1735714

A:Accession: E42645

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-111 <KAU>
A:Cross-references: GB:M80325; NID:9144617; PID:AAA23173.1; PID:9144622

A:Note: sequence extracted from NCBI backbone (NCBIN:79464, NCBI:79469)
R:Stephens, R.S.; Kaiman, S.; Lammell, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,
Science 282, 754-759, 1998

A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trac
A:Reference number: A71570; MUID:9900809; PMID:9784136

A:Accession: C71506

A:Molecule type: DNA

A:Residues: 1-84,'R',86-104,'S',106-111 <ARN>
A:Cross-references: GB:AE00123; GB:AE001273; NID:G3328931; PIDN:AA68118.1; PID:G332895

A:Experimental source: serotype D, strain UW-3/Cx
C:Genetics:

A:Gene: r124
C:Superfamily: Escherichia coli ribosomal protein L24

Query Match 1.9%; Score 8; DB 2; Length 111;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 DKRVVEGI 369
DB 32 DKVVEGI 39

RESULT 10

ribosomal protein L24 TC0804 [imported] - Chlamydia muridarum (strain Nigg)

C:Species: Chlamydia muridarum, Chlamydia trachomatis Mopn
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Jun-2000

C:Accession: D81664

R:Reed, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.
A:Reference number: A81500; MUID:20150255; PMID:10684935

A:Accession: D81664

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-111 <TET>

A:Cross-references: GB:AE002347; GB:AE002160; NID:G7190815; PIDN:AAF39607.1; PID:G719083

C:Genetics:

A:Experimental source: strain Nigg (Mopn)
A:Gene: TC0804
C:Superfamily: Escherichia coli ribosomal protein L24

Query Match 1.9%; Score 8; DB 2; Length 111;

Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 DKRVVEGI 369
DB 32 DKRVVEGI 39

RESULT 11

D70730

hypothetical protein Rv2272 - Mycobacterium tuberculosis (strain H37Rv)

C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: D70730

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
Rajandream, M.A.; Rogers, J.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.
Nature 393, 537-544, 1998

A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: D70730

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-122 <CDL>
A:Cross-references: GB:277163; GB:AL123456; NID:G3261610; PIDN:CAB00963.1; PID:e255069;

A:Experimental source: strain H37Rv
C:Genetics:

A:Gene: Rv2272

Query Match 1.9%; Score 8; DB 2; Length 122;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 LAAVAVLV 21
DB 37 LAAVAVLV 44

RESULT 12

3-oxoacyl-acyl-carrier protein reductase homolog yusR - Bacillus subtilis

C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000

C:Accession: C70022

R:Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azavedo, V.; Bexter
C.; Bron, S.; Brunnelle, S.; Brusch, C.V.; Caldwell, B.; Capiano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi, A.; Galler
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinot,
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauele
V.M.; Ogawa, K.; Ogilwa, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rooha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.M.; Portetelle
A:Authors: Schleich, S.; Schroeder, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot
keuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tosmoni, A.; Tosaro, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamanaka, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
A:Authors: Yoshikawa, H.; Zunshtein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033; PMID:9384377

A:Accession: C70022

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-129 <KUN>
A:Cross-references: GB:Z99120; GB:AL009126; NID:G2635613; PIDN:CAB15279.1; PID:G2635786

A:Experimental source: strain 168
C:Genetics:

A:Gene: yusR

C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

F.1-71/Domains: short-chain alcohol dehydrogenase homology #status atypical <SADH>

Query Match 1.9%; Score 8; DB 2; Length 129;

Best Local Similarity 100.0%; Pred. No. 7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 370 SIAGITGA 377
|||||||
Db 28 SIAGITGA 35

RESULT 13

D90703

Probable fimbrial-like protein [imported] - Escherichia coli (strain O157:H7, substrain C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001

C:Accession: D90703

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G. Gasevaya, N.; Yasunaga, T.; Kuhara, S.; Shibata, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene A:Reference number: A99625; MUID:21156231; PMID:11258796

A:Accession: D90703

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-171 <HAV>

A:Cross-references: GB:BA000007; PIDN:BA034019.1; PID:913360054; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RMD 0509952

C:Genetics:

A:Gene: EC90596

C:Superfamily: type 1 fimbrial protein

Query Match 1.9%; Score 8; DB 2; Length 171;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 201 L1NAGDTT 208
|||||||
Db 61 L1NAGDTT 68

RESULT 14

G85553

Probable fimbrial protein sfmf [imported] - Escherichia coli (strain O157:H7, substrain C:Species: Escherichia coli

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

C:Accession: G85553

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaeser, J.D.; Rose, D.J.; Mayhew

iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoudis, K.; Apodaca,

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: G85553

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-171 <STO>

A:Cross-references: GB:AE005174; NID:912513434; PIDN:AAG54891.1; GSPDB:GN00145; UWGP:Z06

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: sfmf

C:Superfamily: type 1 fimbrial protein

Query Match 1.9%; Score 8; DB 2; Length 171;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 201 L1NAGDTT 208
|||||||
Db 61 L1NAGDTT 68

RESULT 15

E64785

fimbrial protein homolog sfmf - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002

C:Accession: E64785

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C.

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: E64785

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-171 <BLAT>

A:Cross-references: GB:AE000159; GB:U00096; NID:91786739; PIDN:AACT73636.1; PID:91786746;

A:Experimental source: strain K-12, substrain MG1655

C:Genetics:

A:Gene: sfmf; ybcG

C:Superfamily: type 1 fimbrial protein

Query Match 1.9%; Score 8; DB 2; Length 171;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 201 L1NAGDTT 208
|||||||
Db 61 L1NAGDTT 68

RESULT 16

A70673

hypothetical protein Rv2980 - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: A70673

R:Cole, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: A70673

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-181 <COL>

A:Cross-references: GB:Z63018; GB:AL123456; NID:93261671; PIDN:CAB05432.1; PID:E283361;

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: Rv2980

Query Match 1.9%; Score 8; DB 2; Length 181;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 11 AAALAAV 18
|||||||
Db 16 AAALAAV 23

RESULT 17

A72499

hypothetical protein APE1967 - Aeropyrum pernix (strain K1)

C:Species: Aeropyrum pernix

C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000

C:Accession: A72499

R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Hakiawa, Y.; Jin-no, K.; Takah

awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; K

DNA Res. 6, 83-101, 1999

A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr

A:Reference number: A72450; MUID:99310339; PMID:10382966

A:Accession: A72499

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-183 <KAW>

A:Cross-references: DDBJ:AP000063; NID:95105654; PIDN:BAA80977.1; PID:95105665

A:Experimental source: strain K1

C:Genetics:
A:Gene: APR1967
C:Superfamily: conserved hypothetical protein MTH986

Query Match 1.9%; Score 8; DB 2; Length 183;
Best Local Similarity 100.0%; Pred. No. 9.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ALAAVAL 20
DB 118 ALAAVAL 125

RESULT 18

G90849
hypothetical protein ECS1767 [imported] - Escherichia coli (strain O157:H7, substrain R1
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: G90849
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
G:Saikawa, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: G90849
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-183 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA035190.1; PID:G13361232; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain R1MD 0509952
C:Genetics:

A:Gene: ECS1767

Query Match 1.9%; Score 8; DB 2; Length 183;
Best Local Similarity 100.0%; Pred. No. 9.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 267 LIFADPVV 274
DB 130 LIFADPVV 137

RESULT 19

F85707
unknown protein encoded by prophage CP-9330 [imported] - Escherichia coli (strain O157:H
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: F85707
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.D.; Mayhew
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimmlanta, E.; Potamouotis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: AB5480; MUID:21074935; PMID:11206551
A:Accession: F85707
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-183 <STO>
A:Cross-references: GB:AB005174; NID:G12514992; PIDN:AA056122.1; GSPDB:GN00145; UMGF:Z20
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:

A:Gene: Z2047

Query Match 1.9%; Score 8; DB 2; Length 183;
Best Local Similarity 100.0%; Pred. No. 9.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 267 LIFADPVV 274
DB 130 LIFADPVV 137

RESULT 20

136757

probable secreted lipase - Streptomyces coelicolor

C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T36757
R:Saunders, D.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, July 1999
A:Reference number: Z21613
A:Accession: T36757
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 1-290 <SAU>
A:Cross-references: EMBL:AL096849; PIDN:CAB50950.1; GSPDB:GN00070; SCODEB:SC11.24c
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCODEB:SC11.24c

Query Match 1.9%; Score 8; DB 2; Length 290;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 LAAVALV 21
DB 16 LAAVALV 23

RESULT 21

G95961
probable sugar uptake ABC transporter periplasmic solute-binding protein precursor SMD21
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: G95961
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhmester, J.; Chain, P.; Vorholter, F.J.; Hernan
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1.683-kb pSymb megaplasmid from the N2-fixing endo
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: G95961
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-296 <KUB>
A:Cross-references: GB:AL591985; PIDN:CA049355.1; PID:G15140841; GSPDB:GN00167
A:Experimental source: strain 1021, megaplasmid pSymb
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; LeJaure,
habault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: SMD21377
A:Genome: plasmid

Query Match 1.9%; Score 8; DB 2; Length 296;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 AALAAVA 19
DB 216 AALAAVA 223

RESULT 22

H66740
transcription regulator yjfe [imported] - Lactococcus lactis subsp. lactis (strain IL140
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 06-Jan-2003
C:Accession: H66740
R:Boletín, A.; Winkler, P.; Manger, S.; Jallón, O.; Malarme, K.; Weissenbach, J.; Ehrli
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss
A:Reference number: A86625; MUID:21235186; PMID:11337471

A:Accession: H86740
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-306 <STO>
 A:Cross-references: GB:AE005176; PID:g12723863; PIDN:AAK05026.1; GSPDB:GN00146
 A:Experimental source: strain IL403
 C:Genetics:
 A:Gene: yjfe
 C:Superfamily: Bacillus subtilis probable transcription regulator yjvH

Query Match 1.9%; Score 8; DB 2; Length 306;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 200 TLNAGDT 207
 |||||
 Db 202 TLNAGDT 209

RESULT 23

C70415
 cation efflux system (czcB-like) - Aquifex aeolicus
 C:Species: Aquifex aeolicus
 C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999
 C:Accession: C70415
 R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O.V.

Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
 A:Reference number: A70300; MUID:98198666; PMID:9537320

A:Accession: C70415

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-348 <AOE>

A:Cross-references: GB:AE000735; NID:g2983749; PIDN:AA007317.1; PID:g2983757; GB:AE00065

A:Experimental source: strain V85

C:Genetics:
 A:Gene: czcB2

Query Match 1.9%; Score 8; DB 2; Length 348;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 358 LKGDGVV 365
 |||||
 Db 325 LKGDGVV 332

RESULT 24

AC0961
 probable ATP/GTP-binding protein STY3967 [imported] - Salmonella enterica subsp. enteric

C:Species: Salmonella enterica subsp. enterica serovar Typhi
 A:Note: this species has also been called Salmonella typhi

C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C:Accession: AC0961
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Garra, P.

Nature 413, 848-852, 2001
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov

A:Reference number: AB0502; MUID:21534947; PMID:11677608
 A:Accession: AC0961
 A:Status: preliminary
 A:Molecule type: DNA

A:Residues: 1-408 <PAR>
 A:Cross-references: GB:AL513382; PIDN:CAD03183.1; PID:G16504816; GSPDB:GN00176
 C:Genetics:
 A:Gene: STY3967

C:Superfamily: Escherichia coli hypothetical protein yidR

Query Match 1.9%; Score 8; DB 2; Length 408;
 Best Local Similarity 100.0%; Pred. No. 20;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 47 VGVVTVHP 54
 |||||
 Db 64 VGVVTVHP 71

RESULT 25

B65171

hypothetical protein yidR - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002

C:Accession: B65171

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

.A.; Rose, D.U.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: B65171

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-416 <BLAT>

A:Cross-references: GB:AE000446; GB:U00096; NID:g2367261; PIDN:AA06712.1; PID:g1790124;

A:Experimental source: strain K-12, substrain MG1655

C:Genetics:
 A:Gene: yidR

C:Superfamily: Escherichia coli hypothetical protein yidR

Query Match 1.9%; Score 8; DB 2; Length 416;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 47 VGVVTVHP 54
 |||||
 Db 76 VGVVTVHP 83

RESULT 26

E91207

hypothetical protein ECG4629 [imported] - Escherichia coli (strain O157:H7, substrain RI

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001

C:Accession: E91207

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

Gasawara, N.; Yasunaga, T.; Kuhnara, S.; Shiba, T.; Hatori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: E91207

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-416 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA38052.1; PID:g13364104; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain R1MD 0509952

C:Genetics:
 A:Gene: ECG4629

C:Superfamily: Escherichia coli hypothetical protein yidR

Query Match 1.9%; Score 8; DB 2; Length 416;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 47 VGVVTVHP 54
 |||||
 Db 76 VGVVTVHP 83

RESULT 27

G86053

hypothetical protein yidR [imported] - Escherichia coli (strain O157:H7, substrain EDL93

C:Species: Escherichia coli

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

C:Accession: G86053

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaesner, J.D.; Rose, D.J.; Mayhew
 iller, L.; Grobeck, E.J.; Davis, N.W.; Lam, A.; Dimalanta, E.; Polamoukis, K.; Apodaca,
 Nature 409, 529-533, 2001

A>Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: G86053
 A:Status: preliminary

A:Molecule type: DNA
 A:Residues: 1-416 <STO>

A:Cross-references: GB:AE005174; NID:912518530; PIDN:AA658891.1; GSPDB:GN00145; UMGF:251
 A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:
 A:Gene: yldR

C:Superfamily: *Escherichia coli* hypothetical protein yldR

Query Match 1.9%; Score 8; DB 2; Length 416;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 47 VGVVTVHP 54
 |||||
 DB 76 VGVVTVHP 83

RESULT 28

WMBE3
 UA43 protein - human herpesvirus 1 (strain 17)

C:Species: human herpesvirus 1
 C:Date: 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 16-Jun-2000

C:Accession: G30088
 R:McGeoch, D.J.; Dalrymple, M.A.; Davison, A.J.; DoJan, A.; Frame, M.C.; McNab, D.; Perz

J. Gen. Virol. 69, 1531-1574, 1988

A>Title: The complete DNA sequence of the long unique region in the genome of herpes sim
 A:Reference number: A30083; MUID:88274327; PMID:2839594

A:Accession: G30088
 A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA
 A:Residues: 1-434 <MCG>

A:Cross-references: GB:X14112; NID:G1944536; PIDN:CAA32306.1; PID:G59543; GB:D00317
 C:Genetics:

A:Gene: UL43
 C:Superfamily: herpesvirus UL43 protein

Query Match 1.9%; Score 8; DB 1; Length 434;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 AAATAAAV 18
 |||||
 DB 199 AAATAAAV 206

RESULT 29

methionyl aminopeptidase (EC 3.4.11.18) 2 - human
 N:Alternate names: p67; translation initiation factor eIF-2-associated protein

C:Species: Homo sapiens (man)
 C:Date: 14-Jul-1995 #sequence_revision 23-Aug-1996 #text_change 01-Feb-2002

C:Accession: S52112
 R:Li, X.; Chang, Y.H.

Biochim. Biophys. Acta 1260, 333-336, 1995
 A>Title: Molecular cloning of a human complementary DNA encoding an initiation factor 2-

A:Reference number: S52112; MUID:95178556; PMID:7873610

A:Accession: S52112
 A:Molecule type: mRNA

A:Residues: 1-478 <LIX>
 A:Cross-references: GB:U13261; NID:G687242; PIDN:AA63402.1; PID:G687243

C:Genetics:
 A:Gene: GDB:P67EIF2

A:Cross-references: GDB:512821
 C:Function: catalyzes hydrolysis of amino-terminal methionine from proteins
 C:Superfamily: human methionyl aminopeptidase

C:Keywords: aminopeptidase; cobalt; metalloprotein; protein biosynthesis
 F:251,262,459/Binding site: cobalt 2 (Asp, Asp, Glu) #status predicted
 F:262,331,364,459/Binding site: cobalt 1 (Asp, His, Glu, Glu) #status predicted

Query Match 1.9%; Score 8; DB 1; Length 478;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 203 NAGDTTVL 210
 |||||
 DB 235 NAGDTTVL 242

RESULT 30

methionyl aminopeptidase (EC 3.4.11.18) 2 - rat
 A:Accession: A46702

N:Alternate names: p67; translation initiation factor eIF2-associated protein
 C:Species: Rattus norvegicus (Norway rat)

C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 01-Feb-2002
 C:Accession: A46702

R:Mu, S.; Gupta, S.; Chatterjee, N.; Hilleman, R.E.; Kinzy, T.G.; Denlow, N.D.; Merrick,
 J. Biol. Chem. 268, 10796-10801, 1993

A>Title: Cloning and characterization of complementary DNA encoding the eukaryotic initi
 A:Reference number: A46702; MUID:93266517; PMID:8496145

A:Accession: A46702
 A:Molecule type: mRNA

A:Residues: 1-480 <WUA>
 A:Cross-references: GB:L10652; NID:G204003; PIDN:AAA41111.1; PID:G204004

C:Superfamily: human methionyl aminopeptidase
 C:Keywords: aminopeptidase

Query Match 1.9%; Score 8; DB 2; Length 480;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 203 NAGDTTVL 210
 |||||
 DB 235 NAGDTTVL 242

RESULT 31

hypothetical protein F42D1.3 - *Caenorhabditis elegans*.
 C:Species: *Caenorhabditis elegans*

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
 C:Accession: T22088

R:Lightning, J.
 submitted to the EMBL Data Library, October 1996

A:Reference number: Z19511
 A:Accession: T22088

A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA

A:Residues: 1-489 <WII>
 A:Cross-references: EMBL:Z81081; PIDN:CAB03091.1; GSPDB:GN00028; CESP:F42D1.3

A:Experimental source: clone F42D1
 C:Genetics:

A:Gene: CESP:F42D1.3
 A:Map position: X

A:Intons: 21/3; 101/3; 159/2; 201/2; 228/3; 261/2; 322/2; 350/3; 393/3; 420/2

Query Match 1.9%; Score 8; DB 2; Length 489;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 16 AAVALVLS 23
 |||||
 DB 115 AAVALVLS 122

RESULT 32

AB1070

probable carbon starvation protein [imported] - *Salmonella enterica* subsp. *enterica* sero
 C:Species: *Salmonella enterica* subsp. *enterica* serovar Typhi

A>Note: this species has also been called *Salmonella typhi*
 C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
 C:Accession: AB1070
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Garra, P.
 Nature 413, 848-852, 2001
 A:Author: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov A:Reference number: AB0502; MUID:21534947; PMID:11677608
 A:Accession: AB1070
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-716 <PAR>
 A:Cross-references: GB:AL513382; PIDN:CAD03377.1; PID:g16505647; GSPDB:GN00176
 C:Genetics: cstA
 A:Gene: cstA
 C:Superfamily: carbon starvation protein

Query Match 1.9%; Score 8; DB 2; Length 716;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 16 AAVALVS 23
 |||||
 Db 577 AAVALVS 584

RESULT 33

probable integral membrane export protein - *Streptomyces coelicolor*
 C:Species: *Streptomyces coelicolor*
 C>Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
 C:Accession: T35818
 R:Murphy, L.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A. submitted to the EMBL Data Library, February 1999
 A:Reference number: Z21589
 A:Accession: T35818
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-748 <MUR>
 A:Cross-references: EMBL:AL035569; PIDN:CAB37578.1; GSPDB:GN00070; SCOEDB:SCB9.14
 A:Experimental source: strain A3(2)
 C:Genetics:
 A:Gene: SCOEDB:SCB9.14

Query Match 1.9%; Score 8; DB 2; Length 748;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 13 ALAAVAL 20
 |||||
 Db 666 ALAAVAL 673

RESULT 34

membrane alanyl aminopeptidase (EC 3.4.11.2) [imported] - *Yersinia pestis* (strain CO92)
 C:Species: *Yersinia pestis*
 C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 01-Feb-2002
 C:Accession: AH0172
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B. deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; ll, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, Nature 413, 523-527, 2001
 A:Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.
 A:Reference number: AB0001; MUID:21470413; PMID:11586360
 A:Accession: AH0172
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-871 <KUR>
 A:Cross-references: GB:AL590842; PIDN:CAC90243.1; PID:g15979463; GSPDB:GN00175
 C:Genetics:

A:Gene: pepN
 C:Superfamily: microsomal aminopeptidase
 C:Keywords: aminopeptidase

Query Match 1.9%; Score 8; DB 2; Length 871;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 12 ALAAAVA 19
 |||||
 Db 715 ALAAAVA 722

RESULT 35

DEEC
 acetaldehyde dehydrogenase (acetylating) (EC 1.2.1.10) / alcohol dehydrogenase (EC 1.1.1 N:Alternate names: acetaldehyde/alcohol dehydrogenase; Adh; aldehyde reductase
 N:Contains: pyruvate-formate-lyase deactivase
 C:Species: *Escherichia coli*
 C>Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 01-Mar-2002
 C:Accession: J50406; S14809; I76913; D64871; S23871
 R:Goodlove, P.E.; Cunningham, P.R.; Parker, J.; Clark, D.P.
 Gene 85, 209-214, 1989
 A:Title: Cloning and sequence analysis of the fermentative alcohol-dehydrogenase-encoding A:Reference number: J50406; MUID:90152365; PMID:2693398
 A:Accession: J50406
 A:Molecule type: DNA
 A:Residues: 1-891 <GOO>
 A:Cross-references: GB:M3504; NID:g145205; PIDN:AAA23420.1; PID:g145206
 A:Experimental source: plasmid PH18
 A>Note: residues 2-11 were confirmed by protein sequencing
 R:Kessler, D.; Leibrecht, I.; Knappe, J.
 FEBS Lett. 281, 59-63, 1991

A:Title: Pyruvate-formate-lyase-deactivase and acetyl-CoA reductase activities of *Escher A:Reference number: S14809; MUID:91200315; PMID:2015910
 A:Accession: S14809
 A:Molecule type: DNA
 A:Residues: 1-891 <KES>
 A:Cross-references: EMBL:X59263; NID:g40899; PIDN:CAA1955.1; PID:g40900
 A>Note: only nucleotide sequences flanking the coding region are shown
 A>Note: enzyme assays demonstrated stimulation by ferrous ion when either acetyl-CoA or R:Danchn, A.; Kryn, E.
 Microbiology 141, 959-960, 1995
 A:Title: Filling the gap between hns and adh in *Escherichia coli* K12.
 A:Reference number: I57117; MUID:95291445; PMID:7773397
 A:Accession: I76913
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 849-891 <DAN>
 A:Cross-references: EMBL:X67326; NID:g43077; PIDN:CAA47743.1; PID:g43081
 A>Note: submitted to the EMBL Data Library, July 1992
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co A:Rose, D.D.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of *Escherichia coli* K-12.
 A:Reference number: A64720; MUID:97426617; PMID:9278503
 A:Accession: D64871
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-891 <BAT>
 A:Cross-references: GB:AE00222; GB:U00096; NID:g1787486; PIDN:AACT4323.1; PID:g1787493;
 A:Experimental source: strain K-12, substrain M01655
 C:Genetics:
 A:Gene: adhE
 C:Complex: homomultimer containing more than 40 chains
 C:Function: <ADH>
 A:Description: catalyzes the reduction of acetyl-CoA to enzyme-bound thiohemiacetal usin A:Pathway: anaerobic glucose fermentation
 A>Note: operates only in the absence of nitrate
 C:Function: <ARD>
 A:Description: catalyzes the reduction of the thiohemiacetal to ethanol using NADH A:Pathway: anaerobic glucose fermentation
 A>Note: operates only in the absence of nitrate; generates ethanol and NAD+*

C:Superfamily: acetaldehyde/alcohol dehydrogenase; aldehyde dehydrogenase homology; lact
C:Keywords: alcohol metabolism; coenzyme A; homomultimer; iron; multifunctional enzyme;
F:2-691/Product: acetaldehyde/alcohol dehydrogenase #status predicted <MAT>
F:8-266/Domains: aldehyde dehydrogenase homology #link ARD <ALDD>
F:454-858/Domains: lactaldehyde reductase homology #link ARD <LAR>
F:246/Active site: Cys #link ARD #status predicted

Query Match 1.9%; Score 8; DB 1; Length 891;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 RAALALAA 17
|||||||
DB 36 RAALALAA 43

RESULT 36

DB5704
Hypothetical protein adhe [imported] - Escherichia coli (strain O157:H7, substrain EDL93

C:Species: Escherichia coli

C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

C:Accession: DB5704

R:Perma, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew

Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimantanta, E.; Potamoustis, K.; Apodaca,

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: AB5480; MUID:21074935; PMID:11206551

A:Accession: DB5704

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-891 <STO>

A:Cross-references: GB:AE005174; NID:G12514960; PIDN:AG56096.1; GSPDB:GN00145; UMG:Z20

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: adhe

C:Superfamily: acetaldehyde/alcohol dehydrogenase; aldehyde dehydrogenase homology; lact

Query Match 1.9%; Score 8; DB 2; Length 891;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 RAALALAA 17
|||||||
DB 36 RAALALAA 43

RESULT 37

EC0846
acetaldehyde dehydrogenase (acetylating) (EC 1.2.1.10) / alcohol dehydrogenase (EC 1.1.1

C:Species: Escherichia coli

C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 27-Nov-2001

C:Accession: E90846

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: E90846

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-891 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA035164.1; PID:G13361206; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: ECA1741

C:Superfamily: acetaldehyde/alcohol dehydrogenase; aldehyde dehydrogenase homology; lact

C:Keywords: coenzyme A; oxidoreductase

Query Match 1.9%; Score 8; DB 2; Length 891;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 RAALALAA 17

DB 36 RAALALAA 43
|||||||

RESULT 38

AC0265
alcohol dehydrogenase (EC 1.1.1.1) [imported] - Yersinia pestis (strain CO92)

C:Species: Yersinia pestis

C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 11-Jan-2002

C:Accession: AC0265

R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tlhal, R.W.; Holden, M.T.G.; Prentice, M.B.

deno-Farraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.

Il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,

Nature 413, 523-527, 2001

A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A:Reference number: AB0001; MUID:21470413; PMID:11596360

A:Accession: AC0265

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-891 <KUR>

A:Cross-references: GB:AL590842; PIDN:CAC90987.1; PID:G15980182; GSPDB:GN00175

C:Genetics:

A:Gene: adhe

C:Superfamily: acetaldehyde/alcohol dehydrogenase; aldehyde dehydrogenase homology; lact

C:Keywords: oxidoreductase

Query Match 1.9%; Score 8; DB 2; Length 891;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 RAALALAA 17
|||||||
DB 36 RAALALAA 43

RESULT 39

AE0650
alcohol dehydrogenase (EC 1.1.1.1) [imported] - Salmonella enterica subsp. enterica sero

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A:Note: this species has also been called Salmonella typhi

C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C:Accession: AE0650

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,

th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,

S.; Moulé, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica sero

A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AE0650

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-892 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD08384.1; PID:G16502428; GSPDB:GN00176

C:Genetics:

A:Gene: adh

C:Superfamily: acetaldehyde/alcohol dehydrogenase; aldehyde dehydrogenase homology; lact

C:Keywords: oxidoreductase

Query Match 1.9%; Score 8; DB 2; Length 892;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 RAALALAA 17
|||||||
DB 36 RAALALAA 43

RESULT 40

T01566
Hypothetical protein A_T01566.23 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 04-Mar-2000

C:Accession: T01566
R:Dempsey, S.; Harper, M.
submitted to the EMBL Data Library, July 1997
A:Description: The sequence of A. thaliana TW018A10.
A:Reference number: Z14348
A:Accession: T01566
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-979 <DEM>
A:Cross-references: EMBL:AF013294; NID:g2252848; PID:g2252867
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 4
A:introns: 466/3; 569/3; 649/3; 688/1; 740/3; 877/3
A>Note: A TW018A10.23
C:Superfamily: Arabidopsis thaliana hypothetical protein A_TW018A10.23

Query Match 1.9%; Score 8; DB 2; Length 979;
Best Local Similarity 100.0%; Pred.No. 43;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 12 AALAAVA 19
Db 117 AALAAVA 124

Search completed: September 8, 2003, 14:09:40
Job time : 37 secs

THIS PAGE BLANK (USPTO

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2003, 14:07:27 ; Search time 21 Seconds
(without alignments)
830.099 Million cell updates/sec

Title: US-09-889-756A-2

Perfect score: 412
Sequence: 1 MAFVAFKAMRAAALAAVAAL.....AAPQSGVOTASEAKTASAE 412Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database :

1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCITS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfill1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10	2.4	124	4	US-09-461-325-398
2	9	2.2	447	4	US-09-328-352-7383
3	8	1.9	111	4	US-09-732-210-240
4	8	1.9	111	4	US-09-732-210-240
5	8	1.9	194	4	US-09-461-325-397
6	8	1.9	343	4	US-09-252-991A-24458
7	8	1.9	478	2	US-09-040-799-3
8	8	1.9	478	2	US-09-093-448-1
9	8	1.9	478	3	US-09-093-448-2
10	8	1.9	478	3	US-09-093-448-3
11	8	1.9	478	4	US-09-813-555-1
12	8	1.9	478	4	US-09-813-555-2
13	8	1.9	478	4	US-09-813-555-3
14	15	1.7	19	3	US-09-010-999-11
15	14	1.7	24	4	US-09-227-357-347
16	17	1.7	126	1	US-08-198-452A-675
17	17	1.7	126	1	US-08-312-387B-2
18	17	1.7	126	1	US-08-683-426-2
19	17	1.7	126	1	US-08-683-458-2
20	17	1.7	126	2	US-08-878-360-2
21	17	1.7	126	3	US-08-478-140B-2
22	17	1.7	126	4	US-08-333-412-2
23	17	1.7	126	4	US-09-338-943-2
24	7	1.7	131	1	US-08-664-596B-18
25	7	1.7	132	1	US-08-534-975-4
26	7	1.7	132	1	US-08-954-470-4
27	7	1.7	132	3	US-09-129-855A-4

28	7	1.7	132	3	US-09-247-154-4	Sequence 4, Appli
29	7	1.7	132	4	US-09-480-718-4	Sequence 4, Appli
30	7	1.7	132	4	US-09-610-833-4	Sequence 4, Appli
31	7	1.7	132	4	US-09-129-855A-4	Sequence 4, Appli
32	7	1.7	140	4	US-09-252-991A-18327	Sequence 18327, A
33	7	1.7	169	1	US-08-145-995A-7	Sequence 7, Appli
34	7	1.7	169	2	US-08-451-747-7	Sequence 7, Appli
35	7	1.7	169	3	US-09-134-855-7	Sequence 7, Appli
36	7	1.7	195	4	US-09-328-352-6419	Sequence 6419, Ap
37	7	1.7	200	4	US-09-252-991A-18228	Sequence 18228, A
38	7	1.7	262	4	US-09-252-991A-18946	Sequence 18946, A
39	7	1.7	275	4	US-09-107-532A-4013	Sequence 4013, Ap
40	7	1.7	287	4	US-09-107-532A-5035	Sequence 5035, Ap
41	7	1.7	310	4	US-09-134-001C-5326	Sequence 5326, Ap
42	7	1.7	319	4	US-09-252-991A-32369	Sequence 32369, A
43	7	1.7	330	4	US-09-252-991A-18946	Sequence 18946, A
44	7	1.7	357	4	US-09-252-991A-24119	Sequence 24119, A
45	7	1.7	374	1	US-08-095-726-14	Sequence 14, Appl
46	7	1.7	374	1	US-08-096-623A-14	Sequence 14, Appl
47	7	1.7	377	4	US-09-252-991A-32966	Sequence 32966, A
48	7	1.7	402	4	US-09-134-001C-4138	Sequence 4138, Ap
49	7	1.7	402	4	US-09-328-352-4953	Sequence 4953, Ap
50	7	1.7	409	3	US-09-066-046-31	Sequence 31, Appl
51	7	1.7	409	4	US-09-066-047-19	Sequence 19, Appl
52	7	1.7	409	4	US-09-328-352-6858	Sequence 6858, Ap
53	7	1.7	410	4	US-09-252-991A-30606	Sequence 30606, A
54	7	1.7	414	4	US-09-252-991A-27975	Sequence 27975, A
55	7	1.7	418	1	US-08-615-170-16	Sequence 16, Appl
56	7	1.7	422	4	US-09-252-991A-21608	Sequence 21608, A
57	7	1.7	440	4	US-09-252-991A-22487	Sequence 22487, A
58	7	1.7	458	4	US-09-252-991A-30337	Sequence 30337, A
59	7	1.7	479	4	US-09-252-991A-33608	Sequence 33608, A
60	7	1.7	489	1	US-08-434-702-4	Sequence 4, Appli
61	7	1.7	492	3	US-09-006-632-4	Sequence 4, Appli
62	7	1.7	492	3	US-09-006-632-4	Sequence 4, Appli
63	7	1.7	492	4	US-09-325-274-4	Sequence 4, Appli
64	7	1.7	494	3	US-09-006-632-8	Sequence 8, Appli
65	7	1.7	494	3	US-09-006-632-8	Sequence 8, Appli
66	7	1.7	494	4	US-09-325-274-8	Sequence 8, Appli
67	7	1.7	496	3	US-09-348-443-2	Sequence 2, Appli
68	7	1.7	534	4	US-09-252-991A-30902	Sequence 30902, A
69	7	1.7	538	4	US-09-252-991A-22200	Sequence 22200, A
70	7	1.7	542	4	US-09-252-991A-31968	Sequence 31968, A
71	7	1.7	544	2	US-08-933-376A-4	Sequence 4, Appli
72	7	1.7	561	4	US-09-252-991A-23080	Sequence 23080, A
73	7	1.7	561	4	US-09-252-991A-31034	Sequence 31034, A
74	7	1.7	572	2	US-08-933-376A-2	Sequence 2, Appli
75	7	1.7	574	4	US-09-252-991A-18912	Sequence 18912, A
76	7	1.7	574	4	US-09-252-991A-31651	Sequence 31651, A
77	7	1.7	591	4	US-09-519-232-66	Sequence 66, Appl
78	7	1.7	597	4	US-09-252-991A-23152	Sequence 23152, A
79	7	1.7	610	4	US-09-252-991A-24992	Sequence 24992, A
80	7	1.7	615	4	US-09-996-243-194	Sequence 194, App
81	7	1.7	631	4	US-09-311-626B-16	Sequence 16, Appl
82	7	1.7	674	4	US-09-252-991A-24934	Sequence 24934, A
83	7	1.7	687	4	US-09-252-991A-26187	Sequence 26187, A
84	7	1.7	697	4	US-09-328-352-6212	Sequence 6212, Ap
85	7	1.7	705	4	US-09-252-991A-27442	Sequence 27442, A
86	7	1.7	711	4	US-09-252-991A-22259	Sequence 22259, A
87	7	1.7	738	3	US-08-864-038A-3	Sequence 3, Appli
88	7	1.7	742	4	US-09-252-991A-29239	Sequence 29239, A
89	7	1.7	757	4	US-09-252-991A-28180	Sequence 28180, A
90	7	1.7	801	4	US-09-252-991A-18228	Sequence 18228, A
91	7	1.7	1016	4	US-09-252-991A-18914	Sequence 18914, A
92	7	1.7	1057	3	US-08-931-820-4	Sequence 4, Appli
93	7	1.7	1057	3	US-08-963-825-21	Sequence 21, Appl
94	7	1.7	1078	4	US-09-500-811-21	Sequence 21, Appl
95	7	1.7	1078	4	US-09-570-573-21	Sequence 21, Appl
96	7	1.7	1078	4	US-09-548-608-21	Sequence 21, Appl
97	7	1.7	1144	3	US-08-726-214-6	Sequence 6, Appli
98	7	1.7	1186	1	US-08-485-568A-4	Sequence 4, Appli
99	7	1.7	1186	1	US-08-357-698-6	Sequence 6, Appli
100	7	1.7	1186	2	US-08-590-554A-4	Sequence 4, Appli

101	7	1.7	1186	2	US-09-184-223-4	Sequence 4, Appl	174	6	1.5	35	3	US-08-360-107A-49	Sequence 49, Appl
102	7	1.7	1186	5	PCT-US93-12682-6	Sequence 6, Appl	175	6	1.5	35	3	US-08-484-223B-49	Sequence 49, Appl
103	7	1.7	1206	4	US-09-252-991A-19632	Sequence 19632, A	176	6	1.5	35	3	US-08-484-223B-147	Sequence 147, Appl
104	7	1.7	1212	4	US-09-268-866-2	Sequence 2, Appl	177	6	1.5	35	3	US-08-919-597-49	Sequence 49, Appl
105	7	1.7	1700	4	US-09-252-991A-21763	Sequence 21763, A	178	6	1.5	35	3	US-08-919-597-147	Sequence 147, Appl
106	7	1.7	2284	4	US-09-252-991A-23547	Sequence 23547, A	179	6	1.5	35	3	US-08-475-668A-49	Sequence 49, Appl
107	7	1.7	2281	4	US-09-252-991A-21854	Sequence 21854, A	180	6	1.5	35	3	US-08-475-668A-147	Sequence 147, Appl
108	7	1.7	3079	5	PCT-US94-00198-4	Sequence 4, Appl	181	6	1.5	35	3	US-08-485-551A-49	Sequence 49, Appl
109	6	1.5	11	3	US-08-874-347-11	Sequence 11, Appl	182	6	1.5	35	3	US-08-485-551A-147	Sequence 147, Appl
110	6	1.5	11	3	US-09-093-522-11	Sequence 11, Appl	183	6	1.5	35	3	US-08-471-913A-49	Sequence 49, Appl
111	6	1.5	12	2	US-08-874-347-17	Sequence 17, Appl	184	6	1.5	35	3	US-08-471-913A-147	Sequence 147, Appl
112	6	1.5	12	2	US-09-093-522-17	Sequence 17, Appl	185	6	1.5	35	3	US-08-485-546A-49	Sequence 49, Appl
113	6	1.5	13	5	PCT-US95-04121-14	Sequence 14, Appl	186	6	1.5	35	3	US-08-485-546A-147	Sequence 147, Appl
114	6	1.5	14	1	US-08-232-453A-26	Sequence 26, Appl	187	6	1.5	35	3	US-09-082-279B-14	Sequence 14, Appl
115	6	1.5	16	1	US-08-346-849-21	Sequence 21, Appl	188	6	1.5	35	3	US-09-082-279B-153	Sequence 153, Appl
116	6	1.5	16	2	US-08-293-284A-21	Sequence 21, Appl	189	6	1.5	35	3	US-09-082-279B-155	Sequence 155, Appl
117	6	1.5	16	2	US-08-898-300-21	Sequence 21, Appl	190	6	1.5	35	3	US-09-082-279B-156	Sequence 156, Appl
118	6	1.5	20	3	US-08-504-538A-2	Sequence 2, Appl	191	6	1.5	35	3	US-09-082-279B-157	Sequence 157, Appl
119	6	1.5	20	3	US-09-249-458A-2	Sequence 2, Appl	192	6	1.5	35	3	US-09-082-279B-158	Sequence 158, Appl
120	6	1.5	20	4	US-08-630-052-2	Sequence 2, Appl	193	6	1.5	35	3	US-09-082-279B-159	Sequence 159, Appl
121	6	1.5	20	5	PCT-US95-09307-2	Sequence 2, Appl	194	6	1.5	35	3	US-09-082-279B-160	Sequence 160, Appl
122	6	1.5	23	2	US-08-455-968B-21	Sequence 21, Appl	195	6	1.5	35	3	US-09-082-279B-161	Sequence 161, Appl
123	6	1.5	24	3	US-08-504-538A-7	Sequence 7, Appl	196	6	1.5	35	3	US-09-082-279B-162	Sequence 162, Appl
124	6	1.5	24	3	US-09-249-458A-7	Sequence 7, Appl	197	6	1.5	35	3	US-09-082-279B-163	Sequence 163, Appl
125	6	1.5	24	4	US-08-630-052-7	Sequence 7, Appl	198	6	1.5	35	3	US-09-082-279B-164	Sequence 164, Appl
126	6	1.5	24	5	PCT-US95-09307-7	Sequence 7, Appl	199	6	1.5	35	4	US-08-474-349A-288	Sequence 288, Appl
127	6	1.5	25	2	US-08-934-741A-23	Sequence 23, Appl	200	6	1.5	35	4	US-08-474-349A-289	Sequence 289, Appl
128	6	1.5	25	3	US-09-049-691-58	Sequence 58, Appl	201	6	1.5	35	4	US-08-474-349A-147	Sequence 147, Appl
129	6	1.5	26	3	US-09-208-684-2	Sequence 2, Appl	202	6	1.5	35	4	US-08-474-349A-285	Sequence 285, Appl
130	6	1.5	27	3	US-08-809-397-9	Sequence 9, Appl	203	6	1.5	35	4	US-08-474-349A-286	Sequence 286, Appl
131	6	1.5	27	4	US-09-425-597-9	Sequence 9, Appl	204	6	1.5	35	4	US-08-474-349A-287	Sequence 287, Appl
132	6	1.5	27	5	PCT-US95-12502-9	Sequence 9, Appl	205	6	1.5	35	4	US-08-474-349A-288	Sequence 288, Appl
133	6	1.5	28	3	US-08-486-099-146	Sequence 146, App	206	6	1.5	35	4	US-08-474-349A-290	Sequence 290, Appl
134	6	1.5	28	3	US-08-484-223B-146	Sequence 146, App	207	6	1.5	35	4	US-08-474-349A-291	Sequence 291, Appl
135	6	1.5	28	3	US-08-919-597-146	Sequence 146, App	208	6	1.5	35	4	US-08-474-349A-292	Sequence 292, Appl
136	6	1.5	28	3	US-08-475-668A-146	Sequence 146, App	209	6	1.5	35	4	US-08-474-349A-293	Sequence 293, Appl
137	6	1.5	28	3	US-08-485-551A-146	Sequence 146, App	210	6	1.5	35	4	US-08-474-349A-294	Sequence 294, Appl
138	6	1.5	28	3	US-08-471-913A-146	Sequence 146, App	211	6	1.5	35	4	US-08-474-349A-295	Sequence 295, Appl
139	6	1.5	28	3	US-08-485-264A-146	Sequence 146, App	212	6	1.5	35	4	US-08-474-349A-296	Sequence 296, Appl
140	6	1.5	28	3	US-09-082-279B-36	Sequence 36, Appl	213	6	1.5	35	4	US-09-315-304B-34	Sequence 34, Appl
141	6	1.5	28	3	US-09-082-279B-37	Sequence 37, Appl	214	6	1.5	35	4	US-09-315-304B-153	Sequence 153, Appl
142	6	1.5	28	3	US-09-082-279B-38	Sequence 38, Appl	215	6	1.5	35	4	US-09-315-304B-154	Sequence 154, Appl
143	6	1.5	28	4	US-08-474-349A-146	Sequence 146, App	216	6	1.5	35	4	US-09-315-304B-155	Sequence 155, Appl
144	6	1.5	28	4	US-08-474-349A-318	Sequence 318, App	217	6	1.5	35	4	US-09-315-304B-156	Sequence 156, Appl
145	6	1.5	28	4	US-09-315-304B-36	Sequence 36, Appl	218	6	1.5	35	4	US-09-315-304B-157	Sequence 157, Appl
146	6	1.5	28	4	US-09-315-304B-37	Sequence 37, Appl	219	6	1.5	35	4	US-09-315-304B-158	Sequence 158, Appl
147	6	1.5	28	4	US-09-315-304B-38	Sequence 38, Appl	220	6	1.5	35	4	US-09-315-304B-159	Sequence 159, Appl
148	6	1.5	28	4	US-08-470-896-146	Sequence 146, App	221	6	1.5	35	4	US-09-315-304B-160	Sequence 160, Appl
149	6	1.5	28	4	US-08-485-546A-146	Sequence 146, App	222	6	1.5	35	4	US-09-315-304B-161	Sequence 161, Appl
150	6	1.5	28	4	US-09-834-784-36	Sequence 36, Appl	223	6	1.5	35	4	US-09-315-304B-162	Sequence 162, Appl
151	6	1.5	28	4	US-09-834-784-37	Sequence 37, Appl	224	6	1.5	35	4	US-09-315-304B-163	Sequence 163, Appl
152	6	1.5	28	4	US-09-834-784-38	Sequence 38, Appl	225	6	1.5	35	4	US-09-315-304B-164	Sequence 164, Appl
153	6	1.5	29	1	US-07-746-705A-2	Sequence 2, Appl	226	6	1.5	35	4	US-09-315-304B-165	Sequence 165, Appl
154	6	1.5	29	2	US-08-380-182-1	Sequence 1, Appl	227	6	1.5	35	4	US-08-470-896-49	Sequence 49, Appl
155	6	1.5	29	2	US-09-051-934-53	Sequence 53, Appl	228	6	1.5	35	4	US-08-485-546A-49	Sequence 49, Appl
156	6	1.5	30	3	US-09-051-934-54	Sequence 54, Appl	229	6	1.5	35	4	US-08-485-546A-147	Sequence 147, Appl
157	6	1.5	30	6	5188961-8	Patent No. 5188961	230	6	1.5	35	4	US-09-834-784-34	Sequence 34, Appl
158	6	1.5	34	2	US-08-934-741A-3	Sequence 3, Appl	231	6	1.5	35	4	US-09-834-784-153	Sequence 153, Appl
159	6	1.5	34	3	US-08-486-099-19	Sequence 19, Appl	232	6	1.5	35	4	US-09-834-784-154	Sequence 154, Appl
160	6	1.5	34	3	US-08-360-107A-19	Sequence 19, Appl	233	6	1.5	35	4	US-09-834-784-155	Sequence 155, Appl
161	6	1.5	34	3	US-08-484-223B-19	Sequence 19, Appl	234	6	1.5	35	4	US-09-834-784-156	Sequence 156, Appl
162	6	1.5	34	3	US-08-919-597-16	Sequence 19, Appl	235	6	1.5	35	4	US-09-834-784-157	Sequence 157, Appl
163	6	1.5	34	3	US-08-475-668A-19	Sequence 19, Appl	236	6	1.5	35	4	US-09-834-784-158	Sequence 158, Appl
164	6	1.5	34	3	US-08-485-551A-19	Sequence 19, Appl	237	6	1.5	35	4	US-09-834-784-159	Sequence 159, Appl
165	6	1.5	34	3	US-08-471-913A-19	Sequence 19, Appl	238	6	1.5	35	4	US-09-834-784-160	Sequence 160, Appl
166	6	1.5	34	3	US-08-485-264A-19	Sequence 19, Appl	239	6	1.5	35	4	US-09-834-784-161	Sequence 161, Appl
167	6	1.5	34	4	US-08-474-349A-19	Sequence 19, Appl	240	6	1.5	35	4	US-09-834-784-162	Sequence 162, Appl
168	6	1.5	34	4	US-08-470-896-19	Sequence 19, Appl	241	6	1.5	35	4	US-09-834-784-163	Sequence 163, Appl
169	6	1.5	34	4	US-08-485-546A-19	Sequence 19, Appl	242	6	1.5	35	4	US-09-834-784-164	Sequence 164, Appl
170	6	1.5	35	2	US-08-874-347-8	Sequence 8, Appl	243	6	1.5	35	4	US-09-255-208A-19	Sequence 19, Appl
171	6	1.5	35	3	US-08-486-099-49	Sequence 147, Appl	244	6	1.5	37	4	5188961-4	Patent No. 5188961
172	6	1.5	35	3	US-08-486-099-147	Sequence 147, Appl	245	6	1.5	44	6	US-09-057-762-15	Sequence 15, Appl
173	6	1.5	35	3	US-09-093-522-8	Sequence 8, Appl	246	6	1.5	46	2		

247	6	1.5	46	3	US-08-326-119A-15	Sequence 15, Appl	320	6	1.5	152	1	US-08-318-193-84	Sequence 84, Appl
248	6	1.5	48	1	US-07-796-361A-15	Sequence 15, Appl	321	6	1.5	152	4	US-09-198-452A-710	Sequence 110, App
249	6	1.5	59	1	US-08-650-528-7	Sequence 7, Appl	322	6	1.5	155	4	US-09-071-035-112	Sequence 112, App
250	6	1.5	59	3	US-09-060-584-7	Sequence 7, Appl	323	6	1.5	159	4	US-09-252-991A-28211	Sequence 28211, A
251	6	1.5	59	3	US-09-413-140A-7	Sequence 7, Appl	324	6	1.5	162	4	US-09-732-210-1445	Sequence 1445, Ap
252	6	1.5	60	6	5188961-1	Patent No. 5188961	325	6	1.5	163	4	US-09-252-991A-18617	Sequence 18617, A
253	6	1.5	61	4	US-09-205-258-260	Sequence 260, App	326	6	1.5	163	4	US-09-252-991A-20974	Sequence 20974, A
254	6	1.5	64	4	US-09-687-698-22	Sequence 403, App	327	6	1.5	163	4	US-09-252-991A-20974	Sequence 20974, A
255	6	1.5	65	3	US-08-905-223-403	Sequence 4566, Ap	328	6	1.5	165	4	US-09-732-210-1608	Sequence 1608, Ap
256	6	1.5	66	4	US-09-134-001C-4566	Sequence 100, App	329	6	1.5	166	4	US-09-134-001C-3657	Sequence 3657, Ap
257	6	1.5	70	3	US-08-486-099-100	Sequence 100, App	330	6	1.5	166	4	US-09-252-991A-18554	Sequence 18554, A
258	6	1.5	70	3	US-08-360-107A-91	Sequence 91, Appl	331	6	1.5	166	4	US-09-252-991A-22139	Sequence 22139, Ap
259	6	1.5	70	3	US-08-360-107A-110	Sequence 110, App	332	6	1.5	167	4	US-09-134-001C-2867	Sequence 2867, Ap
260	6	1.5	70	3	US-08-484-223B-100	Sequence 100, App	333	6	1.5	169	4	US-09-328-352-4491	Sequence 4491, Ap
261	6	1.5	70	3	US-08-919-597-100	Sequence 100, App	334	6	1.5	170	2	US-08-483-101-3	Sequence 3, Appl
262	6	1.5	70	3	US-08-475-668A-100	Sequence 100, App	335	6	1.5	172	2	US-08-853-659A-46	Sequence 46, Appl
263	6	1.5	70	3	US-08-485-551A-100	Sequence 100, App	336	6	1.5	172	4	US-08-671-548C-50	Sequence 50, Appl
264	6	1.5	70	3	US-08-471-913A-100	Sequence 100, App	337	6	1.5	175	4	US-09-687-698-24	Sequence 24, Appl
265	6	1.5	70	3	US-08-485-264A-100	Sequence 100, App	338	6	1.5	177	1	US-08-284-393B-7	Sequence 7, Appl
266	6	1.5	70	4	US-08-474-349A-100	Sequence 100, App	339	6	1.5	177	1	US-08-446-908-4	Sequence 4, Appl
267	6	1.5	70	4	US-08-255-208A-36	Sequence 36, Appl	340	6	1.5	177	1	US-08-231-205A-4	Sequence 4, Appl
268	6	1.5	70	4	US-08-470-896-100	Sequence 100, App	341	6	1.5	177	2	US-08-850-910A-18	Sequence 18, Appl
269	6	1.5	70	4	US-08-485-546A-100	Sequence 100, App	342	6	1.5	177	2	US-08-871-161-4	Sequence 4, Appl
270	6	1.5	72	4	US-08-918-428D-4	Sequence 4, Appl	343	6	1.5	177	4	US-08-469-260A-49	Sequence 49, Appl
271	6	1.5	75	3	US-08-301-162-6	Sequence 6, Appl	344	6	1.5	177	4	US-08-671-548C-42	Sequence 42, Appl
272	6	1.5	75	4	US-09-461-240-6	Sequence 6, Appl	345	6	1.5	177	4	US-08-671-548C-44	Sequence 44, Appl
273	6	1.5	75	4	US-09-968-927-6	Sequence 6, Appl	346	6	1.5	177	4	US-08-488-446-49	Sequence 49, Appl
274	6	1.5	77	3	US-09-100-802-3	Sequence 3, Appl	347	6	1.5	177	4	US-08-467-344A-49	Sequence 49, Appl
275	6	1.5	78	4	US-09-732-210-820	Sequence 820, App	348	6	1.5	177	5	PCT-US95-08950-7	Sequence 7, Appl
276	6	1.5	79	4	US-09-198-452A-676	Sequence 676, App	349	6	1.5	179	4	US-08-874-102-41	Sequence 41, Appl
277	6	1.5	85	4	US-09-134-001C-5400	Sequence 5400, Ap	350	6	1.5	179	4	US-08-874-102-41	Sequence 41, Appl
278	6	1.5	85	4	US-09-328-352-5627	Sequence 5627, Ap	351	6	1.5	179	4	US-08-984-919A-41	Sequence 41, Appl
279	6	1.5	85	4	US-09-732-210-1469	Sequence 1469, Ap	352	6	1.5	179	4	US-08-984-919A-41	Sequence 41, Appl
280	6	1.5	86	4	US-08-311-731A-242	Sequence 242, App	353	6	1.5	179	4	US-09-252-991A-22072	Sequence 22072, A
281	6	1.5	92	4	US-09-252-991A-16692	Sequence 16692, A	354	6	1.5	179	4	US-09-107-532A-3794	Sequence 3794, Ap
282	6	1.5	94	3	US-08-946-329A-88	Sequence 88, Appl	355	6	1.5	182	4	US-09-252-991A-27281	Sequence 27281, A
283	6	1.5	96	2	US-08-668-255-7	Sequence 7, Appl	356	6	1.5	184	4	US-09-252-991A-22743	Sequence 22743, A
284	6	1.5	96	2	US-08-668-255-9	Sequence 9, Appl	357	6	1.5	185	4	US-09-252-991A-20685	Sequence 20685, A
285	6	1.5	103	2	US-09-252-991A-27103	Sequence 27103, A	358	6	1.5	185	4	US-09-252-991A-23490	Sequence 23490, A
286	6	1.5	105	2	US-08-668-255-5	Sequence 5, Appl	359	6	1.5	186	1	US-08-026-758-25	Sequence 25, Appl
287	6	1.5	105	4	US-09-252-991A-25169	Sequence 25169, A	360	6	1.5	186	1	US-08-026-758-26	Sequence 26, Appl
288	6	1.5	108	4	US-09-252-991A-19573	Sequence 19573, A	361	6	1.5	186	4	US-09-487-792-15	Sequence 15, Appl
289	6	1.5	108	4	US-09-732-210-776	Sequence 776, App	362	6	1.5	186	4	US-09-908-594-15	Sequence 15, Appl
290	6	1.5	111	4	US-09-732-210-778	Sequence 778, App	363	6	1.5	187	4	US-09-252-991A-26269	Sequence 26269, A
291	6	1.5	115	4	US-09-328-352-4375	Sequence 4375, Ap	364	6	1.5	189	4	US-09-082-358B-8	Sequence 8, Appl
292	6	1.5	121	4	US-09-775-932-14	Sequence 14, Appl	365	6	1.5	189	4	US-08-671-548C-48	Sequence 48, Appl
293	6	1.5	122	4	US-09-732-210-241	Sequence 241, App	366	6	1.5	189	4	US-08-671-548C-52	Sequence 52, Appl
294	6	1.5	122	4	US-09-732-210-585	Sequence 585, App	367	6	1.5	190	1	US-08-106-981-2	Sequence 2, Appl
295	6	1.5	123	4	US-09-149-476-693	Sequence 693, App	368	6	1.5	190	4	US-08-918-428D-3	Sequence 3, Appl
296	6	1.5	126	2	US-08-675-508-20	Sequence 20, Appl	369	6	1.5	190	4	US-09-252-991A-18880	Sequence 18880, A
297	6	1.5	127	4	US-08-311-731A-226	Sequence 226, App	370	6	1.5	190	4	US-08-671-548C-32	Sequence 32, Appl
298	6	1.5	128	4	US-09-775-932-12	Sequence 12, Appl	371	6	1.5	193	4	US-08-671-548C-40	Sequence 40, Appl
299	6	1.5	129	4	US-09-732-210-282	Sequence 282, App	372	6	1.5	193	4	US-09-252-991A-17623	Sequence 17623, A
300	6	1.5	134	4	US-09-252-991A-25959	Sequence 25959, A	373	6	1.5	193	4	US-09-252-991A-29167	Sequence 29167, A
301	6	1.5	136	4	US-09-370-838-123	Sequence 123, App	374	6	1.5	194	4	US-09-071-035-110	Sequence 110, App
302	6	1.5	139	4	US-09-252-991A-24414	Sequence 24414, A	375	6	1.5	194	4	US-09-252-991A-31239	Sequence 31239, A
303	6	1.5	141	4	US-09-732-210-1330	Sequence 1330, Ap	376	6	1.5	196	1	US-08-063-552-7	Sequence 7, Appl
304	6	1.5	143	4	US-09-252-991A-23960	Sequence 23960, A	377	6	1.5	196	4	US-09-134-001C-4732	Sequence 4732, Ap
305	6	1.5	143	4	US-09-328-352-4829	Sequence 4829, Ap	378	6	1.5	196	5	PCT-US93-05704-7	Sequence 7, Appl
306	6	1.5	146	2	US-08-619-708A-4	Sequence 4, Appl	379	6	1.5	197	4	US-09-252-991A-31218	Sequence 31218, A
307	6	1.5	146	2	US-09-353-719-1	PCT-US95-07135-2	380	6	1.5	197	4	US-09-252-991A-23002	Sequence 23002, A
308	6	1.5	148	5	US-08-461-030C-2	Sequence 2, Appl	381	6	1.5	199	4	US-09-252-991A-26549	Sequence 26549, A
309	6	1.5	149	2	US-08-744-138-2	Sequence 2, Appl	382	6	1.5	200	4	US-09-107-532A-5925	Sequence 5925, Ap
310	6	1.5	149	3	US-09-431-480-8	Sequence 8, Appl	383	6	1.5	201	1	US-08-444-083-8	Sequence 8, Appl
311	6	1.5	149	3	US-09-431-480-8	Sequence 8, Appl	384	6	1.5	201	1	US-08-286-504-8	Sequence 8, Appl
312	6	1.5	149	3	US-09-431-480-10	Sequence 10, Appl	385	6	1.5	201	1	US-08-442-745-8	Sequence 8, Appl
313	6	1.5	149	3	US-09-617-302-8	Sequence 8, Appl	386	6	1.5	201	1	US-08-443-129-8	Sequence 8, Appl
314	6	1.5	149	3	US-09-617-302-10	Sequence 10, Appl	387	6	1.5	201	1	US-08-443-129-8	Sequence 8, Appl
315	6	1.5	149	4	US-09-241-376-2	Sequence 2, Appl	388	6	1.5	201	1	US-08-443-130-8	Sequence 8, Appl
316	6	1.5	150	4	US-09-252-991A-17264	Sequence 17264, A	389	6	1.5	201	1	US-08-792-019B-11	Sequence 11, Appl
317	6	1.5	151	4	US-09-252-991A-27007	Sequence 27007, A	390	6	1.5	201	3	US-09-106-182-4	Sequence 4, Appl
318	6	1.5	151	4	US-09-732-210-1297	Sequence 1297, Ap	391	6	1.5	201	3	US-08-988-819-11	Sequence 11, Appl
319	6	1.5	151	6	5229115-1	Patent No. 5229115	392	6	1.5	201	3	US-08-898-911-8	Sequence 8, Appl

393	6	1.5	201	3	US-09-016-534-11	Sequence 11, Appl	466	6	1.5	248	3	US-09-417-090-26	Sequence 26, Appl
394	6	1.5	201	4	US-09-648-183-3	Sequence 3, Appl	467	6	1.5	248	3	US-09-189-060B-2	Sequence 2, Appl
395	6	1.5	201	5	PCT-US95-04467-8	Sequence 8, Appl	468	6	1.5	248	3	US-09-349-607-6	Sequence 6, Appl
396	6	1.5	202	4	US-09-252-991A-23766	Sequence 23766, A	469	6	1.5	248	4	US-08-470-953A-6	Sequence 6, Appl
397	6	1.5	203	4	US-09-323-872A-13	Sequence 13, Appl	470	6	1.5	248	4	US-09-727-578-22	Sequence 22, Appl
398	6	1.5	204	4	US-09-149-476-429	Sequence 429, App	471	6	1.5	248	4	US-09-727-578-24	Sequence 24, Appl
399	6	1.5	204	4	US-09-325-932A-145	Sequence 145, App	472	6	1.5	248	4	US-08-750-155A-3	Sequence 26, Appl
400	6	1.5	204	4	US-09-252-991A-23428	Sequence 23428, A	473	6	1.5	249	3	US-08-975-698A-3	Sequence 3, Appl
401	6	1.5	204	4	US-09-252-991A-30470	Sequence 30470, A	474	6	1.5	249	3	US-08-975-698A-3	Sequence 3, Appl
402	6	1.5	204	4	US-09-072-433-22	Sequence 22, Appl	475	6	1.5	249	3	US-09-417-090-3	Sequence 3, Appl
403	6	1.5	205	4	US-09-252-991A-25068	Sequence 25068, A	476	6	1.5	249	4	US-09-727-578-3	Sequence 3, Appl
404	6	1.5	205	4	US-09-252-991A-28196	Sequence 28196, A	477	6	1.5	250	4	US-09-252-991A-25449	Sequence 25449, A
405	6	1.5	205	4	US-09-795-926-37	Sequence 7, Appl	478	6	1.5	253	4	US-09-149-476-576	Sequence 576, App
406	6	1.5	206	1	US-08-709-912-7	Sequence 9, Appl	479	6	1.5	253	4	US-09-198-452A-385	Sequence 385, App
407	6	1.5	206	2	US-09-047-370-7	Sequence 7, Appl	480	6	1.5	254	4	US-09-252-991A-19511	Sequence 19511, A
408	6	1.5	206	4	US-09-107-532A-3801	Sequence 3801, Ap	481	6	1.5	254	4	US-09-252-991A-25583	Sequence 25583, A
409	6	1.5	207	3	US-08-235-836C-36	Sequence 36, Appl	482	6	1.5	255	2	US-08-310-912A-106	Sequence 106, App
410	6	1.5	207	3	US-09-185-501B-14	Sequence 14, Appl	483	6	1.5	255	2	US-08-841-089-106	Sequence 106, App
411	6	1.5	209	4	US-09-795-926-37	Sequence 37, Appl	484	6	1.5	255	3	PCT-US95-04570-106	Sequence 106, App
412	6	1.5	211	4	US-09-252-991A-29679	Sequence 29679, A	485	6	1.5	255	5	PCT-US95-04589-106	Sequence 106, App
413	6	1.5	212	2	US-08-761-248B-2	Sequence 2, Appl	486	6	1.5	255	2	US-08-658-277-2	Sequence 2, Appl
414	6	1.5	212	2	US-08-761-248B-6	Sequence 6, Appl	487	6	1.5	256	2	US-08-658-277-3	Sequence 3, Appl
415	6	1.5	212	2	US-09-199-637A-235	Sequence 235, App	488	6	1.5	256	2	US-09-107-532A-3657	Sequence 3657, Ap
416	6	1.5	213	1	US-07-930-678-2	Sequence 2, Appl	489	6	1.5	256	4	US-09-328-352-6231	Sequence 6231, App
417	6	1.5	213	3	US-08-943-173-3	Sequence 3, Appl	490	6	1.5	257	4	US-09-068-140A-6	Sequence 6, Appl
418	6	1.5	213	4	US-09-252-991A-21832	Sequence 21832, A	491	6	1.5	258	3	US-09-252-991A-28679	Sequence 28679, A
419	6	1.5	214	4	US-09-257-583-15	Sequence 15, Appl	492	6	1.5	259	4	US-08-245-688-2	Sequence 2, Appl
420	6	1.5	217	4	US-09-252-991A-22573	Sequence 22573, A	493	6	1.5	261	1	US-08-245-688-4	Sequence 4, Appl
421	6	1.5	217	4	US-09-252-991A-27623	Sequence 27623, A	494	6	1.5	261	1	US-08-245-688-6	Sequence 6, Appl
422	6	1.5	218	4	US-09-252-991A-29586	Sequence 29586, A	495	6	1.5	261	1	US-08-245-688-8	Sequence 8, Appl
423	6	1.5	221	1	US-08-698-978-2	Sequence 2, Appl	496	6	1.5	261	1	US-08-245-688-10	Sequence 10, Appl
424	6	1.5	221	4	US-08-470-953A-3	Sequence 3, Appl	497	6	1.5	261	1	US-08-245-688-12	Sequence 12, Appl
425	6	1.5	221	4	US-09-252-991A-19158	Sequence 19158, A	498	6	1.5	261	2	US-07-857-224B-58	Sequence 58, Appl
426	6	1.5	222	4	US-08-035-928-21	Sequence 21, Appl	499	6	1.5	261	3	US-08-768-373-2	Sequence 2, Appl
427	6	1.5	224	4	US-09-107-532A-6161	Sequence 6161, Ap	500	6	1.5	261	3	US-09-252-991A-18958	Sequence 18958, A
428	6	1.5	226	3	US-09-195-286-1	Sequence 1, Appl	501	6	1.5	263	4	US-09-702-705-325	Sequence 325, App
429	6	1.5	226	3	US-09-252-991A-11306	Sequence 11306, A	502	6	1.5	263	4	US-09-736-457-325	Sequence 325, App
430	6	1.5	228	3	US-09-303-064-24	Sequence 24, Appl	503	6	1.5	264	4	US-09-252-991A-18073	Sequence 2, Appl
431	6	1.5	228	4	US-09-086-503-24	Sequence 4, Appl	504	6	1.5	266	1	US-08-202-857-2	Sequence 2, Appl
432	6	1.5	229	3	US-08-750-145A-4	Sequence 4, Appl	505	6	1.5	266	1	US-09-589-287B-19	Sequence 19, Appl
433	6	1.5	229	3	US-08-975-698A-4	Sequence 4, Appl	506	6	1.5	266	4	US-09-879-919-24	Sequence 24, Appl
434	6	1.5	229	4	US-09-417-090-4	Sequence 4, Appl	507	6	1.5	266	4	US-09-588-947A-106	Sequence 106, Appl
435	6	1.5	229	4	US-09-727-578-4	Sequence 4, Appl	508	6	1.5	267	4	US-09-252-991A-36069	Sequence 4, Appl
436	6	1.5	230	4	US-09-252-991A-20272	Sequence 20272, A	509	6	1.5	269	1	US-08-202-857-4	Sequence 4, Appl
437	6	1.5	230	4	US-09-328-352-7924	Sequence 16763, A	510	6	1.5	269	1	US-09-082-649B-82	Sequence 82, Appl
438	6	1.5	231	4	US-09-489-847-176	Sequence 176, App	511	6	1.5	269	4	US-09-252-991A-27504	Sequence 27504, A
439	6	1.5	235	3	US-09-101-135-2	Sequence 2, Appl	512	6	1.5	271	1	US-07-977-630-5	Sequence 5, Appl
440	6	1.5	235	3	US-09-303-064-27	Sequence 27, Appl	513	6	1.5	271	2	US-08-850-880-6	Sequence 6, Appl
441	6	1.5	236	4	US-09-086-503-27	Sequence 27, Appl	514	6	1.5	271	2	US-08-944-916-6	Sequence 6, Appl
442	6	1.5	236	4	US-09-252-991A-22298	Sequence 22298, A	515	6	1.5	271	2	US-08-814-877-6	Sequence 6, Appl
443	6	1.5	240	4	US-09-252-991A-21397	Sequence 21397, A	516	6	1.5	271	2	US-08-760-615-8	Sequence 8, Appl
444	6	1.5	240	4	US-09-795-926-33	Sequence 33, Appl	517	6	1.5	271	3	US-09-131-028A-5	Sequence 5, Appl
445	6	1.5	241	3	US-08-896-933-32	Sequence 32, Appl	518	6	1.5	271	3	US-09-414-436-1	Sequence 7, Appl
446	6	1.5	241	3	US-09-314-235-32	Sequence 32, Appl	519	6	1.5	271	3	US-09-131-028A-7	Sequence 7, Appl
447	6	1.5	243	4	US-09-252-991A-18574	Sequence 18574, A	520	6	1.5	271	4	US-09-272-432A-6	Sequence 6, Appl
448	6	1.5	243	4	US-09-252-991A-28884	Sequence 28884, A	521	6	1.5	271	4	US-09-461-325-139	Sequence 139, App
449	6	1.5	244	2	US-09-090-567-2	Sequence 2, Appl	522	6	1.5	271	4	US-09-328-352-6343	Sequence 6343, Ap
450	6	1.5	244	4	US-09-328-352-5628	Sequence 5628, Ap	523	6	1.5	272	3	US-09-101-135-8	Sequence 8, Appl
451	6	1.5	245	4	US-09-252-991A-17982	Sequence 17982, A	524	6	1.5	273	4	US-09-252-991A-32405	Sequence 32405, A
452	6	1.5	245	4	US-09-252-991A-28029	Sequence 28029, A	525	6	1.5	280	4	US-09-252-991A-2698	Sequence 2698, A
453	6	1.5	246	3	US-09-067-800-8	Sequence 8, Appl	526	6	1.5	280	4	US-09-252-991A-29425	Sequence 29425, A
454	6	1.5	246	3	US-09-349-677-8	Sequence 19, Appl	527	6	1.5	282	4	US-09-107-532A-7012	Sequence 7012, Ap
455	6	1.5	247	3	US-08-501-126-19	Sequence 19, Appl	528	6	1.5	282	4	US-09-386-642-54	Sequence 54, Appl
456	6	1.5	248	3	US-08-750-145A-18	Sequence 18, Appl	529	6	1.5	284	4	US-09-522-714-24	Sequence 24, Appl
457	6	1.5	248	3	US-08-750-145A-20	Sequence 20, Appl	530	6	1.5	284	4	US-09-286-529-1	Sequence 1, Appl
458	6	1.5	248	3	US-08-975-698A-22	Sequence 22, Appl	531	6	1.5	285	4	US-09-496-118B-1	Sequence 2, Appl
459	6	1.5	248	3	US-08-975-698A-22	Sequence 22, Appl	532	6	1.5	285	4	US-09-565-423-2	Sequence 2, Appl
460	6	1.5	248	3	US-08-975-698A-24	Sequence 24, Appl	533	6	1.5	285	4	US-09-879-919-23	Sequence 23, Appl
461	6	1.5	248	3	US-08-975-698A-26	Sequence 26, Appl	534	6	1.5	285	4	US-09-252-991A-18133	Sequence 24, Appl
462	6	1.5	248	3	US-09-067-800-6	Sequence 6, Appl	535	6	1.5	285	4		
463	6	1.5	248	3	US-09-417-090-22	Sequence 22, Appl	536	6	1.5	285	4		
464	6	1.5	248	3	US-09-417-090-22	Sequence 22, Appl	537	6	1.5	285	4		
465	6	1.5	248	3	US-09-417-090-24	Sequence 24, Appl	538	6	1.5	285	4		

539	6	1.5	285	4	US-09-588-947A-2	Sequence 2, Appli	612	6	1.5	332	4	US-09-252-991A-28700	Sequence 28700, A
540	6	1.5	287	2	US-08-901-306-2	Sequence 2, Appli	613	6	1.5	333	2	US-08-896-083-2	Sequence 2, Appli
541	6	1.5	287	3	US-09-180-271-2	Sequence 2, Appli	614	6	1.5	333	3	US-08-896-083-4	Sequence 4, Appli
542	6	1.5	287	4	US-09-252-991A-29951	Sequence 29951, A	615	6	1.5	333	3	US-09-110-910A-2	Sequence 2, Appli
543	6	1.5	288	4	US-09-386-642-13	Sequence 13, Appli	616	6	1.5	333	3	US-09-110-910A-4	Sequence 4, Appli
544	6	1.5	289	4	US-09-386-642-14	Sequence 14, Appli	617	6	1.5	334	2	US-08-484-397A-8	Sequence 8, Appli
545	6	1.5	289	4	US-09-252-991A-17482	Sequence 17482, A	618	6	1.5	334	6	5290680-11	Patent No. 5290680
546	6	1.5	289	4	US-08-311-731A-396	Sequence 396, App	619	6	1.5	335	4	US-09-252-991A-25335	Sequence 25335, A
547	6	1.5	290	4	US-09-328-352-7390	Sequence 7390, Ap	620	6	1.5	336	2	US-08-784-651-4	Sequence 4, Appli
548	6	1.5	291	4	US-09-252-991A-22826	Sequence 22826, A	621	6	1.5	336	2	US-08-997-080-156	Sequence 156, App
549	6	1.5	294	4	US-09-252-991A-37760	Sequence 37760, A	622	6	1.5	336	2	US-08-997-362-156	Sequence 156, App
550	6	1.5	294	4	US-09-328-352-4575	Sequence 4575, Ap	623	6	1.5	336	3	US-09-095-855-156	Sequence 156, App
551	6	1.5	295	4	US-09-134-001C-3589	Sequence 3589, Ap	624	6	1.5	336	3	US-09-105-390-52	Sequence 52, Appli
552	6	1.5	297	4	US-08-765-907A-3	Sequence 3, Appli	625	6	1.5	336	3	US-09-105-390-60	Sequence 60, Appli
553	6	1.5	298	4	US-09-252-991A-27658	Sequence 27658, A	626	6	1.5	337	1	US-08-317-223-3	Sequence 3, Appli
554	6	1.5	299	3	US-09-147-826B-2	Sequence 2, Appli	627	6	1.5	337	1	US-08-445-135-4	Sequence 4, Appli
555	6	1.5	299	4	US-09-252-991A-30704	Sequence 30704, A	628	6	1.5	337	3	US-09-055-849A-3	Sequence 3, Appli
556	6	1.5	300	4	US-09-328-352-5297	Sequence 5297, Ap	629	6	1.5	337	3	US-09-213-632-3	Sequence 3, Appli
557	6	1.5	301	4	US-09-252-991A-17405	Sequence 17405, A	630	6	1.5	337	4	US-09-199-637A-315	Sequence 315, App
558	6	1.5	302	4	US-08-887-534A-90	Sequence 90, Appli	631	6	1.5	337	4	US-09-252-991A-21400	Sequence 21400, A
559	6	1.5	302	4	US-09-527-431-90	Sequence 90, Appli	632	6	1.5	337	4	US-09-328-352-5787	Sequence 5787, Ap
560	6	1.5	303	4	US-09-134-001C-3177	Sequence 3177, Ap	633	6	1.5	337	5	PCT-US95-12675-3	Sequence 3, Appli
561	6	1.5	303	4	US-09-674-741-8	Sequence 8, Appli	634	6	1.5	338	4	US-09-328-352-7151	Sequence 7151, Ap
562	6	1.5	303	4	US-09-107-532A-4487	Sequence 4487, Ap	635	6	1.5	340	3	US-09-180-783-2	Sequence 2, Appli
563	6	1.5	304	4	US-09-252-991A-23116	Sequence 23116, A	636	6	1.5	342	4	US-09-252-991A-33584	Sequence 23584, A
564	6	1.5	306	3	US-09-105-390-36	Sequence 36, Appli	637	6	1.5	342	4	US-09-252-991A-27223	Sequence 27223, A
565	6	1.5	306	4	US-09-386-642-53	Sequence 53, Appli	638	6	1.5	343	2	US-08-282-197C-57	Sequence 57, Appli
566	6	1.5	306	4	US-09-252-991A-17554	Sequence 21754, A	639	6	1.5	344	4	US-09-724-623-79	Sequence 79, Appli
567	6	1.5	308	4	US-09-252-991A-18422	Sequence 18422, A	640	6	1.5	344	4	US-09-252-991A-16756	Sequence 16756, A
568	6	1.5	309	1	US-08-236-918A-2	Sequence 2, Appli	641	6	1.5	344	4	US-09-252-991A-19273	Sequence 19273, A
569	6	1.5	309	4	US-09-150-864A-2	Sequence 2, Appli	642	6	1.5	346	4	US-08-366-953A-45	Sequence 45, Appli
570	6	1.5	311	3	US-09-105-390-44	Sequence 44, Appli	643	6	1.5	348	1	US-08-484-397A-2	Sequence 2, Appli
571	6	1.5	312	4	US-09-252-991A-22406	Sequence 22406, A	644	6	1.5	348	2	US-08-484-397A-3	Sequence 3, Appli
572	6	1.5	312	4	US-09-252-991A-31029	Sequence 31029, A	645	6	1.5	348	2	US-08-484-397A-4	Sequence 4, Appli
573	6	1.5	313	2	US-08-284-465-5	Sequence 5, Appli	646	6	1.5	348	2	US-08-484-397A-5	Sequence 5, Appli
574	6	1.5	313	2	US-08-770-544-10	Sequence 10, Appli	647	6	1.5	348	2	US-08-484-397A-6	Sequence 6, Appli
575	6	1.5	313	4	US-09-252-991A-32017	Sequence 32017, A	648	6	1.5	348	2	US-08-484-397A-7	Sequence 7, Appli
576	6	1.5	313	4	US-09-551-826D-14	Sequence 14, Appli	649	6	1.5	348	2	US-08-484-397A-27	Sequence 27, Appli
577	6	1.5	313	4	US-09-579-259-10	Sequence 10, Appli	650	6	1.5	348	2	US-08-484-397A-38	Sequence 38, Appli
578	6	1.5	314	4	US-09-328-352-5492	Sequence 5492, Ap	651	6	1.5	348	2	US-09-300-971A-9	Sequence 9, Appli
579	6	1.5	315	4	US-09-386-653A-9	Sequence 9, Appli	652	6	1.5	351	4	US-09-252-991A-18476	Sequence 18476, A
580	6	1.5	315	4	US-09-252-991A-20012	Sequence 20012, A	653	6	1.5	352	4	US-09-252-991A-18590	Sequence 18590, A
581	6	1.5	316	4	US-09-387-375-9	Sequence 9, Appli	654	6	1.5	353	3	US-08-313-412-1	Sequence 1, Appli
582	6	1.5	317	3	US-08-949-155-6	Sequence 6, Appli	655	6	1.5	353	3	US-09-425-991A-28730	Sequence 28730, A
583	6	1.5	317	4	US-09-819-964-6	Sequence 6, Appli	656	6	1.5	353	4	US-09-328-352-8199	Sequence 8199, Ap
584	6	1.5	317	4	US-09-252-991A-18820	Sequence 18820, A	657	6	1.5	355	4	US-09-252-991A-33034	Sequence 33034, A
585	6	1.5	318	3	US-08-680-506-3	Sequence 3, Appli	658	6	1.5	357	4	US-09-252-991A-32552	Sequence 32552, A
586	6	1.5	318	4	US-09-328-352-6669	Sequence 6669, Ap	659	6	1.5	359	4	US-09-266-965-120	Sequence 120, App
587	6	1.5	319	4	US-09-386-642-12	Sequence 12, Appli	660	6	1.5	359	4	US-09-252-991A-23051	Sequence 23051, A
588	6	1.5	320	4	US-09-526-400-2	Sequence 2, Appli	661	6	1.5	361	1	US-08-457-342-3	Sequence 3, Appli
589	6	1.5	320	4	US-09-252-991A-26003	Sequence 26003, A	662	6	1.5	361	1	US-08-457-646A-3	Sequence 3, Appli
590	6	1.5	321	4	US-09-795-926-35	Sequence 35, Appli	663	6	1.5	361	1	US-08-457-646A-3	Sequence 3, Appli
591	6	1.5	322	4	US-09-252-991A-18724	Sequence 18724, A	664	6	1.5	361	1	US-08-457-646A-3	Sequence 3, Appli
592	6	1.5	323	4	US-09-252-991A-17111	Sequence 17111, A	665	6	1.5	361	1	US-08-457-646A-3	Sequence 3, Appli
593	6	1.5	324	4	US-09-252-991A-27080	Sequence 27080, A	666	6	1.5	361	1	US-08-457-646A-3	Sequence 3, Appli
594	6	1.5	325	4	US-08-611-107-12	Sequence 12, Appli	667	6	1.5	361	1	US-08-457-646A-3	Sequence 3, Appli
595	6	1.5	327	2	US-08-422-560A-12	Sequence 12, Appli	668	6	1.5	361	1	US-08-457-646A-3	Sequence 3, Appli
596	6	1.5	327	2	US-09-173-581-5	Sequence 5, Appli	669	6	1.5	361	2	US-08-729-214-3	Sequence 3, Appli
597	6	1.5	327	3	US-08-468-793-12	Sequence 12, Appli	670	6	1.5	361	3	US-09-028-934-30	Sequence 30, Appli
598	6	1.5	327	3	US-09-420-915-5	Sequence 5, Appli	671	6	1.5	361	3	US-09-501-115-8	Sequence 8, Appli
599	6	1.5	327	3	US-09-386-642-11	Sequence 11, Appli	672	6	1.5	361	3	US-09-252-991A-26917	Sequence 26917, A
600	6	1.5	327	3	US-09-252-991A-17509	Sequence 17509, A	673	6	1.5	361	3	US-09-252-991A-20195	Sequence 20195, A
601	6	1.5	327	3	US-09-252-991A-17509	Sequence 17509, A	674	6	1.5	361	3	US-09-252-991A-423	Sequence 423, App
602	6	1.5	327	3	US-09-252-991A-17509	Sequence 17509, A	675	6	1.5	361	3	US-08-984-919A-11	Sequence 11, Appli
603	6	1.5	327	3	US-09-252-991A-17509	Sequence 17509, A	676	6	1.5	361	3	US-08-545-573A-1	Sequence 1, Appli
604	6	1.5	328	4	US-09-252-991A-17729	Sequence 17729, A	677	6	1.5	366	4	US-09-795-926-39	Sequence 39, Appli
605	6	1.5	328	4	US-09-252-991A-17729	Sequence 17729, A	678	6	1.5	367	4	US-09-252-991A-19846	Sequence 19846, A
606	6	1.5	330	4	US-09-134-001C-3582	Sequence 3582, Ap	679	6	1.5	367	4	US-08-781-420-11	Sequence 11, Appli
607	6	1.5	330	4	US-09-134-001C-3582	Sequence 3582, Ap	680	6	1.5	368	3	US-08-874-102-11	Sequence 11, Appli
608	6	1.5	331	4	US-09-134-001C-3582	Sequence 3582, Ap	681	6	1.5	368	3	US-09-006-555A-11	Sequence 11, Appli
609	6	1.5	331	4	US-09-134-001C-3582	Sequence 3582, Ap	682	6	1.5	368	3	US-09-006-555A-11	Sequence 11, Appli
610	6	1.5	331	4	US-09-134-001C-3582	Sequence 3582, Ap	683	6	1.5	368	3	US-09-006-555A-11	Sequence 11, Appli
611	6	1.5	331	4	US-09-134-001C-3582	Sequence 3582, Ap	684	6	1.5	368	3	US-09-006-555A-11	Sequence 11, Appli

685	6	1.5	369	1	US-08-700-359-21	Sequence 21, Appl	758	6	1.5	405	4	US-09-252-991A-2388	Sequence 2388, A
686	6	1.5	369	4	US-09-252-991A-17960	Sequence 17960, A	759	6	1.5	406	4	US-09-198-452A-387	Sequence 387, App
687	6	1.5	369	4	US-08-311-731A-38	Sequence 38, Appl	760	6	1.5	409	4	US-09-613-303-55	Sequence 55, Appl
688	6	1.5	370	4	US-09-252-991A-21152	Sequence 21152, A	761	6	1.5	409	4	US-09-328-352-450	Sequence 450, Ap
689	6	1.5	370	4	US-08-311-731A-203	Sequence 203, App	762	6	1.5	410	4	US-09-252-991A-3050	Sequence 3050, A
690	6	1.5	372	4	US-09-107-532A-4779	Sequence 4779, Ap	763	6	1.5	412	4	US-09-252-991A-22952	Sequence 22952, A
691	6	1.5	373	3	US-08-746-883-4	Sequence 4, Appl	764	6	1.5	414	3	US-09-334-601-4	Sequence 4, Appl
692	6	1.5	373	3	US-09-188-579-113	Sequence 113, App	765	6	1.5	414	4	US-09-252-991A-24714	Sequence 24714, A
693	6	1.5	373	3	US-09-315-444-113	Sequence 113, App	766	6	1.5	417	4	US-09-252-991A-18947	Sequence 18947, A
694	6	1.5	373	3	US-09-721-362-113	Sequence 113, App	767	6	1.5	416	4	US-09-599-360B-88	Sequence 88, Appl
695	6	1.5	375	1	US-08-121-714-5	Sequence 5, Appl	768	6	1.5	418	4	US-09-252-991A-17826	Sequence 17826, A
696	6	1.5	375	1	US-08-205-719-2	Sequence 2, Appl	769	6	1.5	419	3	US-09-011-197-4	Sequence 4, Appl
697	6	1.5	375	1	US-08-477-108A-5	Sequence 5, Appl	770	6	1.5	420	4	US-09-252-991A-30920	Sequence 30920, A
698	6	1.5	375	2	US-08-477-112-5	Sequence 5, Appl	771	6	1.5	421	4	US-09-198-452A-932	Sequence 932, App
699	6	1.5	375	2	US-08-746-883-5	Sequence 5, Appl	772	6	1.5	423	2	US-08-715-554-2	Sequence 2, Appl
700	6	1.5	375	5	PCT-US93-08322-5	Sequence 5, Appl	773	6	1.5	423	2	US-08-583-118-2	Sequence 2, Appl
701	6	1.5	376	4	US-09-252-991A-20454	Sequence 20454, A	774	6	1.5	423	4	US-09-072-009-142	Sequence 142, App
702	6	1.5	377	2	US-08-455-968B-3	Sequence 3, Appl	775	6	1.5	424	4	US-09-252-991A-17882	Sequence 17882, A
703	6	1.5	377	4	US-09-410-464-8	Sequence 8, Appl	776	6	1.5	424	4	US-09-328-352-4199	Sequence 4199, Ap
704	6	1.5	377	4	US-09-252-991A-19827	Sequence 19827, A	777	6	1.5	426	4	US-09-327-487A-1	Sequence 1, Appl
705	6	1.5	378	2	US-08-455-968B-10	Sequence 10, Appl	778	6	1.5	426	4	US-09-327-487A-5	Sequence 5, Appl
706	6	1.5	378	2	US-08-823-516-138	Sequence 138, App	779	6	1.5	426	4	US-09-327-487A-6	Sequence 6, Appl
707	6	1.5	378	4	US-09-252-991A-23047	Sequence 23047, A	780	6	1.5	426	4	US-09-327-487A-7	Sequence 7, Appl
708	6	1.5	378	4	US-09-328-352-8178	Sequence 8178, Ap	781	6	1.5	426	4	US-09-071-035-2	Sequence 2, Appl
709	6	1.5	379	2	US-08-887-365-36	Sequence 36, Appl	782	6	1.5	427	4	US-09-134-001C-3801	Sequence 3801, Ap
710	6	1.5	379	4	US-09-252-991A-31643	Sequence 31643, A	783	6	1.5	428	4	US-09-134-001C-3801	Sequence 3801, Ap
711	6	1.5	380	2	US-08-455-968B-1	Sequence 1, Appl	784	6	1.5	428	4	US-09-252-991A-25955	Sequence 25955, A
712	6	1.5	380	2	US-08-823-516-137	Sequence 137, App	785	6	1.5	428	4	US-08-486-039-95	Sequence 95, Appl
713	6	1.5	380	4	US-09-426-557-10	Sequence 10, Appl	786	6	1.5	430	3	US-08-360-107A-105	Sequence 105, App
714	6	1.5	380	4	US-09-252-991A-32470	Sequence 32470, A	787	6	1.5	430	3	US-08-484-223B-95	Sequence 95, Appl
715	6	1.5	380	4	US-09-107-532A-3873	Sequence 3873, Ap	788	6	1.5	430	3	US-08-484-223B-95	Sequence 95, Appl
716	6	1.5	380	5	PCT-US91-02560-4	Sequence 4, Appl	789	6	1.5	430	3	US-08-919-597-95	Sequence 95, Appl
717	6	1.5	382	4	US-09-252-991A-23727	Sequence 23727, A	790	6	1.5	430	3	US-08-475-668A-95	Sequence 95, Appl
718	6	1.5	382	4	US-09-252-991A-30714	Sequence 30714, A	791	6	1.5	430	3	US-08-485-551A-95	Sequence 95, Appl
719	6	1.5	382	4	US-09-328-352-7748	Sequence 7748, Ap	792	6	1.5	430	3	US-08-471-913A-95	Sequence 95, Appl
720	6	1.5	384	3	US-08-968-563-12	Sequence 12, Appl	793	6	1.5	430	3	US-08-485-264A-95	Sequence 95, Appl
721	6	1.5	384	3	US-08-969-683A-12	Sequence 12, Appl	794	6	1.5	430	4	US-08-474-349A-95	Sequence 95, Appl
722	6	1.5	384	4	US-09-297-928-8	Sequence 8, Appl	795	6	1.5	430	4	US-08-255-208A-31	Sequence 31, Appl
723	6	1.5	386	3	US-09-293-322C-5	Sequence 5, Appl	796	6	1.5	430	4	US-08-470-896-95	Sequence 95, Appl
724	6	1.5	386	4	US-09-325-932A-143	Sequence 143, App	797	6	1.5	430	4	US-08-485-546A-95	Sequence 95, Appl
725	6	1.5	386	4	US-08-545-573A-2	Sequence 2, Appl	798	6	1.5	430	4	US-09-252-991A-18400	Sequence 18400, A
726	6	1.5	386	4	US-08-545-573A-39	Sequence 39, Appl	799	6	1.5	431	4	US-09-107-532A-7079	Sequence 7079, Ap
727	6	1.5	386	4	US-09-839-497A-5	Sequence 5, Appl	800	6	1.5	432	4	US-09-328-352-4131	Sequence 4131, Ap
728	6	1.5	386	4	US-09-198-452A-388	Sequence 388, App	801	6	1.5	434	4	US-09-252-991A-30855	Sequence 30855, A
729	6	1.5	387	4	US-09-252-991A-18477	Sequence 18477, A	802	6	1.5	435	4	US-09-252-991A-27532	Sequence 27532, A
730	6	1.5	387	4	US-09-252-991A-22112	Sequence 22112, A	803	6	1.5	435	3	US-09-252-991A-18163	Sequence 18163, A
731	6	1.5	390	4	US-09-634-238-331	Sequence 331, App	804	6	1.5	436	3	US-09-315-444-112	Sequence 112, App
732	6	1.5	392	4	US-09-647-224A-14	Sequence 14, Appl	805	6	1.5	436	4	US-09-721-362-112	Sequence 112, App
733	6	1.5	392	4	US-09-252-991A-29764	Sequence 29764, A	806	6	1.5	437	4	US-09-252-991A-24390	Sequence 24390, A
734	6	1.5	395	3	US-09-032-372-1	Sequence 1, Appl	807	6	1.5	438	4	US-09-198-452A-985	Sequence 985, App
735	6	1.5	396	2	US-08-878-989-16	Sequence 16, Appl	808	6	1.5	439	4	US-09-198-452A-986	Sequence 986, App
736	6	1.5	396	2	US-08-850-880-4	Sequence 4, Appl	809	6	1.5	442	4	US-09-252-991A-31743	Sequence 31743, A
737	6	1.5	396	2	US-08-944-916-4	Sequence 4, Appl	810	6	1.5	443	4	US-09-252-991A-18213	Sequence 18213, A
738	6	1.5	396	2	US-08-814-877-4	Sequence 4, Appl	811	6	1.5	443	4	US-09-252-991A-30036	Sequence 30036, A
739	6	1.5	396	3	US-09-108-020-38	Sequence 38, Appl	812	6	1.5	444	4	US-09-252-991A-32552	Sequence 32552, A
740	6	1.5	396	3	US-09-272-796-16	Sequence 16, Appl	813	6	1.5	444	4	US-09-252-991A-17767	Sequence 17767, A
741	6	1.5	396	3	US-09-344-700-2	Sequence 2, Appl	814	6	1.5	444	4	US-09-252-991A-24777	Sequence 24777, A
742	6	1.5	396	4	US-09-272-432A-4	Sequence 4, Appl	815	6	1.5	446	4	US-09-252-991A-21819	Sequence 21819, A
743	6	1.5	397	1	US-08-415-823-2	Sequence 2, Appl	816	6	1.5	448	4	US-09-252-991A-30230	Sequence 30230, A
744	6	1.5	397	2	US-09-086-662-2	Sequence 2, Appl	817	6	1.5	449	3	US-08-680-506-7	Sequence 7, Appl
745	6	1.5	399	4	US-08-311-731A-250	Sequence 250, App	818	6	1.5	449	4	US-09-252-991A-17164	Sequence 17164, A
746	6	1.5	398	2	US-08-599-171A-29	Sequence 29, Appl	819	6	1.5	450	4	US-09-252-991A-26266	Sequence 26266, A
747	6	1.5	398	2	US-08-646-590B-29	Sequence 29, Appl	820	6	1.5	450	4	US-09-369-247-97	Sequence 97, Appl
748	6	1.5	398	3	US-09-069-226-29	Sequence 29, Appl	821	6	1.5	451	4	US-09-252-991A-31986	Sequence 31986, A
749	6	1.5	398	3	US-09-412-184-29	Sequence 29, Appl	822	6	1.5	454	4	US-09-252-991A-38633	Sequence 38633, A
750	6	1.5	399	4	US-09-252-991A-16821	Sequence 16821, A	823	6	1.5	455	4	US-09-252-991A-18636	Sequence 18636, A
751	6	1.5	399	4	US-09-252-991A-20042	Sequence 20042, A	824	6	1.5	456	2	US-08-643-034A-4	Sequence 4, Appl
752	6	1.5	399	4	US-09-252-991A-23741	Sequence 23741, A	825	6	1.5	456	3	US-08-648-650A-4	Sequence 4, Appl
753	6	1.5	400	4	US-09-252-991A-24985	Sequence 24985, A	826	6	1.5	456	4	US-09-465-558-50	Sequence 50, Appl
754	6	1.5	401	4	US-09-252-991A-21672	Sequence 21672, A	827	6	1.5	457	3	US-09-431-517F-14	Sequence 14, Appl
755	6	1.5	403	2	US-08-846-762-89	Sequence 89, Appl	828	6	1.5	457	4	US-09-183-268A-4	Sequence 4, Appl
756	6	1.5	404	4	US-09-252-991A-24408	Sequence 24408, A	829	6	1.5	457	4	US-09-252-991A-19877	Sequence 19877, A
757	6	1.5	405	4	US-09-252-991A-23282	Sequence 23282, A	830	6	1.5	457	4	US-09-252-991A-22570	Sequence 22570, A

831	6	1.5	457	4	US-08-311-731A-145	Sequence 145, App	904	6	1.5	501	3	US-09-307-185-6	Sequence 6, Appli
832	6	1.5	458	4	US-09-252-991A-20811	Sequence 20811, A	905	6	1.5	501	4	US-09-157-257-8	Sequence 8, Appli
833	6	1.5	458	4	US-09-252-991A-28585	Sequence 28585, A	906	6	1.5	501	4	US-09-252-991A-17730	Sequence 17730, A
834	6	1.5	460	4	US-09-252-991A-27768	Sequence 27768, A	907	6	1.5	502	4	US-09-252-991A-24388	Sequence 24388, A
835	6	1.5	460	4	US-09-252-991A-32153	Sequence 32153, A	908	6	1.5	503	4	US-09-252-991A-18052	Sequence 18052, A
836	6	1.5	461	1	US-07-796-361A-13	Sequence 13, Appli	909	6	1.5	503	4	US-09-252-991A-29343	Sequence 29343, A
837	6	1.5	462	1	US-07-612-673-4	Sequence 4, Appli	910	6	1.5	506	3	US-09-303-064-5-2	Sequence 52, Appli
838	6	1.5	462	1	US-08-539-666-4	Sequence 4, Appli	911	6	1.5	506	3	US-09-086-503-12	Sequence 52, Appli
839	6	1.5	462	4	US-09-252-991A-31372	Sequence 31372, A	912	6	1.5	506	4	US-09-252-991A-19050	Sequence 19050, A
840	6	1.5	462	4	US-09-328-352-7128	Sequence 7128, Ap	913	6	1.5	506	4	US-09-252-991A-19150	Sequence 19150, A
841	6	1.5	463	4	US-09-252-991A-30002	Sequence 30002, A	914	6	1.5	506	4	US-09-252-991A-32150	Sequence 32150, A
842	6	1.5	464	4	US-09-252-991A-26602	Sequence 26602, A	915	6	1.5	507	4	US-09-130-337A-25	Sequence 25, Appli
843	6	1.5	466	4	US-08-984-919A-33	Sequence 33, Appli	916	6	1.5	507	4	US-09-252-991A-23598	Sequence 23598, A
844	6	1.5	467	4	US-09-252-991A-29057	Sequence 29057, A	917	6	1.5	508	4	US-09-369-247-167	Sequence 167, App
845	6	1.5	468	4	US-08-874-102-33	Sequence 33, Appli	918	6	1.5	509	4	US-09-252-991A-32556	Sequence 32556, A
846	6	1.5	469	3	US-08-378-313-33	Sequence 33, Appli	919	6	1.5	510	3	US-08-246-489-2	Sequence 2, Appli
847	6	1.5	470	4	US-08-984-919A-55	Sequence 55, Appli	920	6	1.5	510	4	US-09-252-991A-25076	Sequence 25076, A
848	6	1.5	470	4	US-09-252-991A-20500	Sequence 20500, A	921	6	1.5	512	4	US-09-390-234-24	Sequence 24, Appli
849	6	1.5	471	4	US-09-252-991A-27633	Sequence 27633, A	922	6	1.5	512	4	US-09-716-865-14	Sequence 14, Appli
850	6	1.5	472	4	US-08-874-102-55	Sequence 55, Appli	923	6	1.5	514	4	US-09-252-991A-22124	Sequence 22124, A
851	6	1.5	472	4	US-09-252-991A-20558	Sequence 20558, A	924	6	1.5	516	2	US-08-762-106-8	Sequence 8, Appli
852	6	1.5	472	4	US-09-252-991A-22840	Sequence 22840, A	925	6	1.5	516	2	US-08-676-166A-2	Sequence 2, Appli
853	6	1.5	474	3	US-09-315-444-116	Sequence 116, App	926	6	1.5	516	3	US-08-745-404-2	Sequence 2, Appli
854	6	1.5	474	4	US-09-721-362-116	Sequence 116, App	927	6	1.5	516	3	US-09-320-774-8	Sequence 8, Appli
855	6	1.5	474	4	US-09-252-991A-21280	Sequence 21280, A	928	6	1.5	519	4	US-09-252-991A-30815	Sequence 30815, A
856	6	1.5	476	3	US-09-188-579-114	Sequence 114, App	929	6	1.5	520	4	US-09-252-991A-33049	Sequence 33049, A
857	6	1.5	476	3	US-09-315-444-114	Sequence 114, App	930	6	1.5	521	4	US-09-134-001C-4290	Sequence 4290, Ap
858	6	1.5	476	4	US-09-134-218-6	Sequence 6, Appli	931	6	1.5	521	4	US-09-252-991A-30925	Sequence 30925, A
859	6	1.5	476	4	US-09-721-362-114	Sequence 114, App	932	6	1.5	522	4	US-08-469-260A-606	Sequence 606, App
860	6	1.5	476	4	US-09-529-157-4	Sequence 4, Appli	933	6	1.5	522	4	US-08-468-446-606	Sequence 606, App
861	6	1.5	480	1	US-07-882-292-2	Sequence 2, Appli	934	6	1.5	522	4	US-08-467-344A-606	Sequence 606, App
862	6	1.5	480	2	US-08-468-812-5	Sequence 5, Appli	935	6	1.5	523	4	US-09-252-991A-23875	Sequence 23875, A
863	6	1.5	480	2	US-08-331-644-2	Sequence 2, Appli	936	6	1.5	523	4	US-09-252-991A-33001	Sequence 33001, A
864	6	1.5	480	4	US-08-590-563-5	Sequence 5, Appli	937	6	1.5	526	4	US-09-252-991A-28402	Sequence 28402, A
865	6	1.5	480	4	US-09-770-621-5	Sequence 5, Appli	938	6	1.5	526	4	US-09-252-991A-32984	Sequence 32984, A
866	6	1.5	480	5	PCT-US93-04102-2	Sequence 13, Appli	939	6	1.5	527	2	US-08-762-106-9	Sequence 9, Appli
867	6	1.5	483	3	US-08-431-517F-13	Sequence 13, Appli	940	6	1.5	527	3	US-09-320-774-9	Sequence 9, Appli
868	6	1.5	485	2	US-09-252-991A-29267	Sequence 29267, A	941	6	1.5	529	4	US-09-252-991A-18188	Sequence 18188, A
869	6	1.5	485	2	US-08-749-391-2	Sequence 2, Appli	942	6	1.5	529	4	US-09-252-991A-18188	Sequence 18188, A
870	6	1.5	485	3	US-09-390-200-2	Sequence 2, Appli	943	6	1.5	530	4	US-09-252-991A-21863	Sequence 21863, A
871	6	1.5	486	3	US-09-906-743-4	Sequence 4, Appli	944	6	1.5	531	4	US-09-252-991A-21965	Sequence 21965, A
872	6	1.5	487	4	US-09-594-193-13	Sequence 13, Appli	945	6	1.5	531	4	US-09-252-991A-22509	Sequence 22509, A
873	6	1.5	488	4	US-09-252-991A-19034	Sequence 19034, A	946	6	1.5	531	4	US-09-252-991A-31109	Sequence 31109, A
874	6	1.5	488	4	US-09-252-991A-24316	Sequence 24316, A	947	6	1.5	532	1	US-08-339-152A-32	Sequence 32, Appli
875	6	1.5	488	4	US-09-252-991A-29894	Sequence 29894, A	948	6	1.5	532	1	US-08-657-192-9	Sequence 9, Appli
876	6	1.5	489	1	US-09-252-991A-25740	Sequence 25740, A	949	6	1.5	532	3	US-08-523-373-7	Sequence 7, Appli
877	6	1.5	490	1	US-08-201-118-5	Sequence 5, Appli	950	6	1.5	532	3	US-08-948-564-10	Sequence 10, Appli
878	6	1.5	490	1	US-08-201-118-11	Sequence 11, Appli	951	6	1.5	533	4	US-09-252-991A-22953	Sequence 22953, A
879	6	1.5	490	2	US-08-238-821B-5	Sequence 5, Appli	952	6	1.5	534	4	US-09-252-991A-17757	Sequence 17757, A
880	6	1.5	490	2	US-08-238-821B-11	Sequence 11, Appli	953	6	1.5	535	4	US-09-252-991A-24728	Sequence 24728, A
881	6	1.5	490	5	PCT-US95-05744-5	Sequence 5, Appli	954	6	1.5	535	4	US-09-252-991A-25662	Sequence 25662, A
882	6	1.5	490	5	PCT-US95-05744-11	Sequence 11, Appli	955	6	1.5	537	1	US-08-657-192-15	Sequence 15, Appli
883	6	1.5	492	2	US-08-468-812-4	Sequence 4, Appli	956	6	1.5	539	2	US-08-467-963C-2	Sequence 2, Appli
884	6	1.5	492	2	US-08-468-812-7	Sequence 7, Appli	957	6	1.5	539	3	US-08-838-189D-2	Sequence 2, Appli
885	6	1.5	492	4	US-08-590-563-4	Sequence 4, Appli	958	6	1.5	539	3	US-08-852-344D-2	Sequence 2, Appli
886	6	1.5	492	4	US-08-590-563-7	Sequence 7, Appli	959	6	1.5	539	3	US-08-344-639E-2	Sequence 2, Appli
887	6	1.5	492	4	US-09-770-621-4	Sequence 4, Appli	960	6	1.5	539	3	US-08-467-969A-2	Sequence 2, Appli
888	6	1.5	492	4	US-09-770-621-7	Sequence 7, Appli	961	6	1.5	539	3	US-08-467-961A-2	Sequence 2, Appli
889	6	1.5	493	3	US-09-177-349-5	Sequence 5, Appli	962	6	1.5	539	3	US-08-001-554A-2	Sequence 6, Appli
890	6	1.5	493	4	US-09-252-991A-30722	Sequence 30722, A	963	6	1.5	539	4	US-09-157-257-6	Sequence 6, Appli
891	6	1.5	493	4	US-09-252-991A-32463	Sequence 32463, A	964	6	1.5	539	4	US-09-328-352-5433	Sequence 5433, Ap
892	6	1.5	494	4	US-09-252-991A-24099	Sequence 24099, A	965	6	1.5	540	4	US-09-252-991A-22219	Sequence 22219, A
893	6	1.5	494	4	US-09-252-991A-32739	Sequence 32739, A	966	6	1.5	540	4	US-09-328-352-4514	Sequence 4514, Ap
894	6	1.5	495	4	US-08-984-919A-47	Sequence 47, Appli	967	6	1.5	541	2	US-08-540-804-16	Sequence 16, Appli
895	6	1.5	495	4	US-09-252-991A-16768	Sequence 16768, A	968	6	1.5	541	2	US-08-218-265-16	Sequence 16, Appli
896	6	1.5	497	4	US-08-874-102-47	Sequence 47, Appli	969	6	1.5	541	3	US-08-521-872-1	Sequence 1, Appli
897	6	1.5	498	2	US-08-511-485-13	Sequence 13, Appli	970	6	1.5	541	3	US-08-590-399-16	Sequence 16, Appli
898	6	1.5	498	4	US-09-134-001C-5233	Sequence 5233, Ap	971	6	1.5	542	4	US-09-252-991A-31091	Sequence 31091, A
899	6	1.5	498	4	US-09-201-936-13	Sequence 13, Appli	972	6	1.5	543	4	US-09-252-991A-20430	Sequence 20430, A
900	6	1.5	498	4	US-09-252-991A-17608	Sequence 17608, A	973	6	1.5	545	4	US-09-252-991A-17065	Sequence 17065, A
901	6	1.5	498	4	US-09-252-991A-20702	Sequence 20702, A	974	6	1.5	547	4	US-09-252-991A-33044	Sequence 33044, A
902	6	1.5	499	4	US-09-252-991A-23929	Sequence 23929, A	975	6	1.5	551	1	US-08-120-960-2	Sequence 2, Appli
903	6	1.5	501	2	US-08-980-060-6	Sequence 6, Appli	976	6	1.5	551	2	US-08-700-548-4	Sequence 4, Appli

```

977 6 1.5 551 4 US-09-347-878-9 Sequence 9, Appli
978 6 1.5 552 3 US-08-745-404-3 Sequence 3, Appli
979 6 1.5 553 2 US-08-663-566A-13 Sequence 13, Appli
980 6 1.5 553 2 US-08-484-575A-14 Sequence 14, Appli
981 6 1.5 553 2 US-08-023-610-13 Sequence 13, Appli
982 6 1.5 553 2 US-08-288-065A-13 Sequence 13, Appli
983 6 1.5 553 2 US-08-362-240A-13 Sequence 13, Appli
984 6 1.5 553 2 US-08-477-459-14 Sequence 14, Appli
985 6 1.5 553 3 US-08-479-869-14 Sequence 14, Appli
986 6 1.5 553 3 US-08-486-414-14 Sequence 14, Appli
987 6 1.5 553 3 US-08-804-372A-11 Sequence 11, Appli
988 6 1.5 553 4 US-09-252-991A-17984 Sequence 17984, A
989 6 1.5 553 4 US-09-252-991A-19864 Sequence 19864, A
990 6 1.5 553 5 PCT-US94-01826A-14 Sequence 14, Appli
991 6 1.5 553 5 PCT-US94-02252A-14 Sequence 14, Appli
992 6 1.5 553 5 PCT-US95-10245-13 Sequence 13, Appli
993 6 1.5 553 6 5310678-1 Patent No. 5310678
994 6 1.5 554 4 US-09-252-991A-23106 Sequence 23106, A
995 6 1.5 555 4 US-09-252-991A-28313 Sequence 28313, A
996 6 1.5 555 4 US-09-252-991A-28848 Sequence 28848, A
997 6 1.5 556 4 US-09-252-991A-24832 Sequence 24832, A
998 6 1.5 557 2 US-08-808-982-6 Sequence 6, Appli
999 6 1.5 557 3 US-09-306-902A-6 Sequence 6, Appli
1000 6 1.5 558 4 US-09-328-352-5580 Sequence 5580, Ap

```

ALIGNMENTS

```

RESULT 1
US-09-461-325-398
; Sequence 398, Application US/09461325A
; Patent No. 6475753
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 94 Human Secreted Proteins
; FILE REFERENCE: P2029PI
; CURRENT APPLICATION NUMBER: US/09/461,325A
; CURRENT FILING DATE: 1999-12-14
; EARLIER APPLICATION NUMBER: PCT/US99/13418
; EARLIER FILING DATE: 1999-06-15
; EARLIER APPLICATION NUMBER: 60/089,507
; EARLIER FILING DATE: 1998-06-16
; EARLIER APPLICATION NUMBER: 60/089,508
; EARLIER FILING DATE: 1998-06-16
; EARLIER APPLICATION NUMBER: 60/089,509
; EARLIER FILING DATE: 1998-06-16
; EARLIER APPLICATION NUMBER: 60/089,510
; EARLIER FILING DATE: 1998-06-16
; EARLIER APPLICATION NUMBER: 60/090,112
; EARLIER FILING DATE: 1998-06-22
; EARLIER APPLICATION NUMBER: 60/090,113
; EARLIER FILING DATE: 1998-06-22
; NUMBER OF SEQ ID NOS: 532
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 398
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-461-325-398

Query Match          2.4%; Score 10; DB 4; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.044;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7383
; LENGTH: 447
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7383

Query Match          2.2%; Score 9; DB 4; Length 447;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

RESULT 3
US-09-732-210-240
; Sequence 240, Application US/09732210
; Patent No. 6573361
; GENERAL INFORMATION:
; APPLICANT: Bunkers, Greg J.
; APPLICANT: Liang, Jihong
; APPLICANT: Mitcanck, Cindy A.
; APPLICANT: Seale, Jeffrey W.
; APPLICANT: Wu, Yonnie S.
; TITLE OF INVENTION: Anti-fungal Proteins and Methods for Their Use
; FILE REFERENCE: 38-21(15036)B
; CURRENT APPLICATION NUMBER: US/09/732,210
; CURRENT FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/169,513
; PRIOR FILING DATE: 1999-12-07
; PRIOR APPLICATION NUMBER: US 60/169,340
; PRIOR FILING DATE: 1999-12-07
; NUMBER OF SEQ ID NOS: 1753
; SEQ ID NO 240
; LENGTH: 111
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis
US-09-732-210-240

```

```

Query Match          1.9%; Score 8; DB 4; Length 111;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY          362 DKVVVEGI 369
Db          32 DKVVVEGI 39

RESULT 4
US-09-732-210-764
; Sequence 764, Application US/09732210
; Patent No. 6573361
; GENERAL INFORMATION:
; APPLICANT: Bunkers, Greg J.
; APPLICANT: Liang, Jihong
; APPLICANT: Mitcanck, Cindy A.
; APPLICANT: Seale, Jeffrey W.
; APPLICANT: Wu, Yonnie S.
; TITLE OF INVENTION: Anti-fungal Proteins and Methods for Their Use
; FILE REFERENCE: 38-21(15036)B
; CURRENT APPLICATION NUMBER: US/09/732,210
; CURRENT FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/169,513

```


;; PRIOR FILING DATE: 1999-12-07
;; PRIOR APPLICATION NUMBER: US 60/169,340
;; PRIOR FILING DATE: 1999-12-07
;; NUMBER OF SEQ ID NOS: 1753
;; SEQ ID NO 764
;; LENGTH: 111
;; TYPE: PRT
;; ORGANISM: Chlamydia trachomatis
US-09-732-210-764

Query Match 1.9%; Score 8; DB 4; Length 111;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 362 DKVVEGI 369
Db 32 DKVVEGI 39

RESULT 5
US-09-461-325-397
; Sequence 397, Application US/09461325A
; Patent No. 6475753
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 94 Human Secreted Proteins
; FILE REFERENCE: P2029P1
; CURRENT APPLICATION NUMBER: US/09/461,325A
; EARLIER APPLICATION NUMBER: PCT/US99/13418
; EARLIER FILING DATE: 1999-06-15
; EARLIER APPLICATION NUMBER: 60/089,507
; EARLIER FILING DATE: 1998-06-16
; EARLIER APPLICATION NUMBER: 60/089,508
; EARLIER FILING DATE: 1998-06-16
; EARLIER APPLICATION NUMBER: 60/089,509
; EARLIER FILING DATE: 1998-06-16
; EARLIER APPLICATION NUMBER: 60/089,510
; EARLIER FILING DATE: 1998-06-16
; EARLIER APPLICATION NUMBER: 60/090,112
; EARLIER FILING DATE: 1998-06-22
; EARLIER APPLICATION NUMBER: 60/090,113
; EARLIER FILING DATE: 1998-06-22
; NUMBER OF SEQ ID NOS: 532
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 397
; LENGTH: 194
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-461-325-397

Query Match 1.9%; Score 8; DB 4; Length 194;
Best Local Similarity 100.0%; Pred. No. 6.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 AAAAAVA 19
Db 176 AAAAAVA 183

RESULT 6
US-09-252-991A-24458
; Sequence 24458, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18

;; PRIOR APPLICATION NUMBER: US 60/094,190
;; PRIOR FILING DATE: 1998-07-27
;; NUMBER OF SEQ ID NOS: 33142
;; SEQ ID NO 24458
;; LENGTH: 343
;; TYPE: PRT
;; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24458

Query Match 1.9%; Score 8; DB 4; Length 343;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 AAAAAAV 18
Db 225 AAAAAAV 232

RESULT 7
US-09-040-799-3
; Sequence 3, Application US/09040799
; Patent No. 585820
; GENERAL INFORMATION:
; APPLICANT: CHANG, YIE-HWA
; TITLE OF INVENTION: CLONE OF A NUCLEOTIDE SEQUENCE ENCODING
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAFERKAMP, L.C.
; STREET: 7733 FORSYTH BLVD., SUITE 1400
; CITY: ST. LOUIS
; STATE: MO
; COUNTRY: USA
; ZIP: 63105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040,799
; FILING DATE: 18-MAR-1998
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 16153-4639
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 314-727-6092
; TELEFAX: 314-727-5188
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 478 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-040-799-3

Query Match 1.9%; Score 8; DB 2; Length 478;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 203 NAGDTTVL 210
Db 235 NAGDTTVL 242

RESULT 8
US-09-093-448-1
; Sequence 1, Application US/09093448A
; Patent No. 6207704
; GENERAL INFORMATION:

```

; APPLICANT: Liu, Jun O.
; APPLICANT: Griffith, Eric C.
; APPLICANT: Su, Zhuang
; TITLE OF INVENTION: Type 2 Methionine Aminopeptidase (MetAP2) Inhibitors
; FILE REFERENCE: 0492611-0346
; CURRENT APPLICATION NUMBER: US/09/093,448A
; CURRENT FILING DATE: 1998-06-08
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-093-448-1

Query Match          1.9%; Score 8; DB 3; Length 478;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      203 NAGDTTVL 210
DB      235 NAGDTTVL 242

RESULT 9
US-09-093-448-2
; Sequence 2, Application US/09093448A
; Patent No. 6207704
; GENERAL INFORMATION:
; APPLICANT: Liu, Jun O.
; APPLICANT: Griffith, Eric C.
; APPLICANT: Su, Zhuang
; TITLE OF INVENTION: Type 2 Methionine Aminopeptidase (MetAP2) Inhibitors
; FILE REFERENCE: 0492611-0346
; CURRENT APPLICATION NUMBER: US/09/093,448A
; CURRENT FILING DATE: 1998-06-08
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-093-448-2

Query Match          1.9%; Score 8; DB 3; Length 478;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      203 NAGDTTVL 210
DB      235 NAGDTTVL 242

RESULT 10
US-09-093-448-3
; Sequence 3, Application US/09093448A
; Patent No. 6207704
; GENERAL INFORMATION:
; APPLICANT: Liu, Jun O.
; APPLICANT: Griffith, Eric C.
; APPLICANT: Su, Zhuang
; TITLE OF INVENTION: Type 2 Methionine Aminopeptidase (MetAP2) Inhibitors
; FILE REFERENCE: 0492611-0346
; CURRENT APPLICATION NUMBER: US/09/093,448A
; CURRENT FILING DATE: 1998-06-08
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 478
; TYPE: PRT
```

```

; ORGANISM: Homo sapiens
US-09-093-448-3

Query Match          1.9%; Score 8; DB 3; Length 478;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      203 NAGDTTVL 210
DB      235 NAGDTTVL 242

RESULT 11
US-09-813-555-1
; Sequence 1, Application US/09813555
; Patent No. 6566541
; GENERAL INFORMATION:
; APPLICANT: Liu, Jun O.
; APPLICANT: Griffith, Eric C.
; APPLICANT: Su, Zhuang
; TITLE OF INVENTION: Type 2 Methionine Aminopeptidase (MetAP2) Inhibitors
; FILE REFERENCE: 0492611-0346
; CURRENT APPLICATION NUMBER: US/09/813,555
; CURRENT FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-813-555-1

Query Match          1.9%; Score 8; DB 4; Length 478;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      203 NAGDTTVL 210
DB      235 NAGDTTVL 242

RESULT 12
US-09-813-555-2
; Sequence 2, Application US/09813555
; Patent No. 6566541
; GENERAL INFORMATION:
; APPLICANT: Liu, Jun O.
; APPLICANT: Griffith, Eric C.
; APPLICANT: Su, Zhuang
; TITLE OF INVENTION: Type 2 Methionine Aminopeptidase (MetAP2) Inhibitors
; FILE REFERENCE: 0492611-0346
; CURRENT APPLICATION NUMBER: US/09/813,555
; CURRENT FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-813-555-2

Query Match          1.9%; Score 8; DB 4; Length 478;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      203 NAGDTTVL 210
DB      235 NAGDTTVL 242

RESULT 13
```

US-09-813-555-3
Sequence 3, Application US/09813555
Patent No. 6566541
GENERAL INFORMATION:
APPLICANT: Liu, Jun O.
APPLICANT: Griffith, Eric C.
TITLE OF INVENTION: Type 2 Methionine Aminopeptidase (MetAP2) Inhibitors
TITLE OF INVENTION: and Uses Thereof
FILE REFERENCE: 0492611-0346
CURRENT APPLICATION NUMBER: US/09/813,555
CURRENT FILING DATE: 2001-03-21
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3
LENGTH: 478
TYPE: PRT
ORGANISM: Homo sapiens
US-09-813-555-3

Query Match 1.9%; Score 8; DB 4; Length 478;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 203 NAGDTTVL 210
Db 235 NAGDTTVL 242

RESULT 14
US-09-010-999-11

Sequence 11, Application US/09010999
Patent No. 6132976

GENERAL INFORMATION:

APPLICANT: Poole, Anthony R.
APPLICANT: Hollander, Anthony P.
APPLICANT: Billingham, R. C.

TITLE OF INVENTION: IMMUNOASSAYS FOR THE MEASUREMENT OF
TITLE OF INVENTION: COLLAGEN DENATURATION AND CLEAVAGE IN CARTILAGE

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington

STATE: D.C.
COUNTRY: USA

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/010,999

FILING DATE: 22-JAN-1998

CLASSIFICATION: 4335

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/448,501

FILING DATE: 17-JUL-1995

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/984,123

FILING DATE: 04-DEC-1992

ATTORNEY/AGENT INFORMATION:
NAME: Bent, Stephen A.

REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 032931/0212

TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399

TELEX: 904136
INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:
LENGTH: 19 amino acids

TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-010-999-11

Query Match 1.7%; Score 7; DB 3; Length 19;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 371 IAGITGA 377
Db 11 IAGITGA 17

RESULT 15

US-09-227-357-347
Sequence 347, Application US/09227357
Patent No. 6342581

GENERAL INFORMATION:

APPLICANT: Fischer et al.
TITLE OF INVENTION: 123 Human Secreted Proteins

FILE REFERENCE: P2010P1
CURRENT APPLICATION NUMBER: US/09/227,357

FILING DATE: 1999-01-08
EARLIER APPLICATION NUMBER: PCT/US98/13684

FILING DATE: 1998-07-07
EARLIER APPLICATION NUMBER: 60/051,926

FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,793

FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,925

FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,929

FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,803

FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,732

FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,931

FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,932

FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,916

FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,930

FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,918

FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,920

FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,733

FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,795

FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,919

FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,928

FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/055,722

FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,723

FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,948

FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,949

FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,953

FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,950

FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,947

FILING DATE: 1997-08-18

```

; EARLIER APPLICATION NUMBER: 60/055,964
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/056,360
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,684
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,984
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,954
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/058,785
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,664
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,660
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,661
; EARLIER FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 347
; LENGTH: 24
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-227-357-347

```

```

Query Match      1.7%; Score 7; DB 4; Length 24;
Best Local Similarity 100.0%; Pred. No. 8.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      111 NLESARA 117
      |||||
DB      8 NLESARA 14

```

```

RESULT 16
US-09-198-452A-675
; Sequence 675, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Grifffais, R
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 675
; LENGTH: 111
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-675

```

```

Query Match      1.7%; Score 7; DB 4; Length 111;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      362 DKVVEG 368
      |||||
DB      32 DKVVEG 38

```

```

RESULT 17
US-08-312-387B-2
; Sequence 2, Application US/08312387B
; Patent No. 5545553
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSER: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/312,387B
; FILING DATE: July 7, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 126 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-312-387B-2

```

```

Query Match      1.7%; Score 7; DB 1; Length 126;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      11 AAAA1AA 17
      |||||
DB      13 AAAA1AA 19

```

```

RESULT 18
US-08-683-426-2
; Sequence 2, Application US/08683426
; Patent No. 5705367
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/683,426
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/312,387
; FILING DATE: September 26, 1994
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095B

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ. ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-683-458-2

Query Match 1.7%; Score 7; DB 1; Length 126;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 11 AAALAAA 17
13 AAALAAA 19

RESULT 19
US-08-683-458-2
Sequence 2, Application US/08683458
Patent No. 5798233

GENERAL INFORMATION:
APPLICANT: Gotschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/683,458
FILING DATE:

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/312,387
FILING DATE: September 26, 1994

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-095A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521

INFORMATION FOR SEQ. ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-683-458-2

Query Match 1.7%; Score 7; DB 1; Length 126;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 11 AAALAAA 17
13 AAALAAA 19

RESULT 20
US-08-878-360-2
Sequence 2, Application US/08878360
Patent No. 5945322

GENERAL INFORMATION:
APPLICANT: Gotschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/878,360
FILING DATE: 18-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/683,426
FILING DATE:

APPLICATION NUMBER: 08/312,387
FILING DATE: September 26, 1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-095B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521

INFORMATION FOR SEQ. ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-878-360-2

Query Match 1.7%; Score 7; DB 2; Length 126;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 11 AAALAAA 17
13 AAALAAA 19

RESULT 21
US-08-478-140B-2
Sequence 2, Application US/08478140B
Patent No. 6127153

GENERAL INFORMATION:
APPLICANT: JOHNSON, KARL F.
APPLICANT: ROTH, STEPHEN
APPLICANT: BUCZALA, STEPHANIE L.
TITLE OF INVENTION: METHOD OF TRANSFERRING AT LEAST TWO
TITLE OF INVENTION: SACCHARIDE UNITS WITH A POLYGLYCOSYLTRANSFERASE, A
TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE AND GENE ENCODING A
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds

STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,140B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Laura A. Coruzzi
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7188-017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-478-140B-2

Query Match 1.7%; Score 7; DB 3; Length 126;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 AAALAA 17
Db 13 AAALAA 19

RESULT 22
US-09-333-412-2
Sequence 2, Application US/09333412
Patent No. 6342382
GENERAL INFORMATION:
APPLICANT: Gotschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
OLIGOSACCHARIDES, AND GENES ENCODING THEM
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/333,412
FILING DATE: 15-Jun-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/312,387
FILING DATE: July 7, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-095
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800

TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-333-412-2

Query Match 1.7%; Score 7; DB 4; Length 126;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 AAALAA 17
Db 13 AAALAA 19

RESULT 23
US-09-338-943-2
Sequence 2, Application US/09338943
Patent No. 6379933
GENERAL INFORMATION:
APPLICANT: JOHNSON, KARL F.
APPLICANT: ROTH, STEPHEN
APPLICANT: BUCZALA, STEPHANIE L.
TITLE OF INVENTION: METHOD OF TRANSFERRING AT LEAST TWO
SACCHARIDE UNITS WITH A POLYGLYCOSYLTRANSFERASE, A
TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE AND GENE ENCODING A
TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/338,943
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/478,140
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Laura A. Coruzzi
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7188-017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-338-943-2

Query Match 1.7%; Score 7; DB 4; Length 126;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 AAALAA 17

Db 13 AAALAA 19

RESULT 24
US-08-664-596B-18

Sequence 18, Application US/08664596B

Patent No. 5807703

GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth

APPLICANT: McCoy, John

APPLICANT: Lavallie, Edward

APPLICANT: Racie, Lisa

APPLICANT: Treacy, David

APPLICANT: Treacy, Maurice

APPLICANT: Evans, Cheryl

APPLICANT: Spaulding, Vikki

APPLICANT: Bowman, Michael

TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES

TITLE OF INVENTION: ENCODING THEM

NUMBER OF SEQUENCES: 37

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/664,596B

FILING DATE:

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Brown, Scott A.

REGISTRATION NUMBER: 32,724

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 498-8224

TELEFAX: (617) 498-5851

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 131 amino acids

TYPE: amino acid

STRANDNESS:

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-664-596B-18

Query Match 1.7%; Score 7; DB 1; Length 131;

Best Local Similarity 100.0%; Pred. No. 46;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 403 SEAKTAS 409
Db 97 SEAKTAS 103

RESULT 25

US-08-534-975-4

Sequence 4, Application US/08534975

Patent No. 5723313

GENERAL INFORMATION:

APPLICANT: Sherr, Charles, J.

APPLICANT: Quelle, Dawn, E.

TITLE OF INVENTION: ARF-p19, A No. 5723313el Regulator of the Mammalian Cell

TITLE OF INVENTION: Cycle

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.

STREET: 1100 New York Avenue Suite 600

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/534,975

FILING DATE: 28-SEP-1995

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Fox, Samuel, L.

REGISTRATION NUMBER: 30,353

REFERENCE/DOCKET NUMBER: 0656.0590000

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-371-2540

TELEFAX: 202-371-2540

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 132 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-534-975-4

Query Match 1.7%; Score 7; DB 1; Length 132;

Best Local Similarity 100.0%; Pred. No. 47;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 AAVALVL 22
Db 41 AAVALVL 47

RESULT 26

US-08-954-470-4

Sequence 4, Application US/08954470

Patent No. 5876965

GENERAL INFORMATION:

APPLICANT: Sherr, Charles, J.

APPLICANT: Quelle, Dawn, E.

TITLE OF INVENTION: ARF-p19, A No. 5876965el Regulator of the

TITLE OF INVENTION: Mammalian Cell Cycle

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.

STREET: 1100 New York Avenue Suite 600

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/954,470

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/534,975

FILING DATE: 28-SEP-1995

ATTORNEY/AGENT INFORMATION:

NAME: Fox, Samuel, L.

REGISTRATION NUMBER: 30,353

REFERENCE/DOCKET NUMBER: 0656.0590000

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-371-2600

TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 132 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-954-470-4

Query Match 1.7%; Score 7; DB 2; Length 132;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 AAVAVL 22
| | | | |
41 AAVAVL 47

Db

RESULT 27
US-09-129-855A-4

Sequence 4, Application US/09129855A
Patent No. 6046032

GENERAL INFORMATION:

APPLICANT: Sherr, Charles, J.

APPLICANT: Quelle, Dawn, E.

TITLE OF INVENTION: ARF-p19, A No. 6046032e1 Regulator of the Mammalian

TITLE OF INVENTION: Cell Cycle

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sterne, Kessler, Goldstein & Fox

STREET: 1100 New York Avenue Suite 600

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/129,855A

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/954,470

FILING DATE: 20-OCT-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/534,975

FILING DATE: 27-SEP-1995

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 1340-1-023 CIP

TELEPHONE: 201-487-5800

TELEFAX: 201-343-1684

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 132 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-129-855A-4

Query Match 1.7%; Score 7; DB 3; Length 132;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 AAVAVL 22
| | | | |
41 AAVAVL 47

Db

RESULT 28

US-09-247-154-4

Sequence 4, Application US/09247154

Patent No. 6172194

GENERAL INFORMATION:

APPLICANT: Sherr, Charles, J.

APPLICANT: Quelle, Dawn, E.

TITLE OF INVENTION: ARF-p19, A No. 6172194e1 Regulator of the

TITLE OF INVENTION: Mammalian Cell Cycle

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.

STREET: 1100 New York Avenue Suite 600

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/247,154

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/954,470

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Fox, Samuel, L.

REGISTRATION NUMBER: 30,353

REFERENCE/DOCKET NUMBER: 0656.0590000

TELEPHONE: 202-371-2600

TELEFAX: 202-371-2540

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 132 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-247-154-4

Query Match 1.7%; Score 7; DB 3; Length 132;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 AAVAVL 22
| | | | |
41 AAVAVL 47

Db

RESULT 29

US-09-480-718-4

Sequence 4, Application US/09480718

Patent No. 6407062

GENERAL INFORMATION:

APPLICANT: Sherr, Charles J.

APPLICANT: Quelle, Dawn E.

APPLICANT: Weber, Jason D.

APPLICANT: Rousset, Martine F.

APPLICANT: Frederique, Zidny

TITLE OF INVENTION: ARF-p19, A NOVEL REGULATOR OF THE MAMMALIAN CELL CYCLE

FILE REFERENCE: 1340-1-023 CIP 1

CURRENT APPLICATION NUMBER: US/09/480,718

CURRENT FILING DATE: 2000-01-07

EARLIER FILING DATE: 1998-08-06

NUMBER OF SEQ ID NOS: 48

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 4


```

; LENGTH: 132
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-480-718-4

Query Match
Best Local Similarity 1.7%; Score 7; DB 4; Length 132;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 16 AAVALVL 22
   |||||
Db 41 AAVALVL 47

RESULT 30
US-09-610-833-4
; Sequence 4, Application US/09610833
; Patent No. 6482929
; GENERAL INFORMATION:
; APPLICANT: Sherr, Charles, J.
; TITLE OF INVENTION: ARR-p19, A No. 6482929el Regulator of the
; Mammalian Cell Cycle
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESS: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/610,833
; FILING DATE: 06-Jul-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/954,470
; FILING DATE: <Unknown>
; APPLICATION NUMBER: 08/534,975
; FILING DATE: 28-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fox, Samuel, L.
; REGISTRATION NUMBER: 30,353
; REFERENCE/DOCKET NUMBER: 0656.0590000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 132 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-610-833-4

Query Match
Best Local Similarity 1.7%; Score 7; DB 4; Length 132;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 16 AAVALVL 22
   |||||
Db 41 AAVALVL 47

RESULT 31
US-09-129-855A-4
; Sequence 4, Application US/09129855A

; Patent No. 6586203
; GENERAL INFORMATION:
; APPLICANT: Sherr, Charles, J.
; TITLE OF INVENTION: ARR-p19, A No. 6586203el Regulator of the Mammalian
; Cell Cycle
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESS: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/129,855A
; FILING DATE: 06-Aug-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/954,470
; FILING DATE: 20-OCT-1997
; APPLICATION NUMBER: 08/534,975
; FILING DATE: 27-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1340-1-023 CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 132 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-129-855A-4

Query Match
Best Local Similarity 1.7%; Score 7; DB 4; Length 132;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 16 AAVALVL 22
   |||||
Db 41 AAVALVL 47

RESULT 32
US-09-252-991A-18327
; Sequence 18327, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18327
; LENGTH: 140
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
```

US-09-252-991A-18327

Query Match 1.7%: Score 7; DB 4; Length 140;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ALAAVA 19
|||||
DB 101 ALAAVA 107

RESULT 33

US-08-145-995A-7
; Sequence 7, Application US/08145995A
; Patent No. 5482850
; GENERAL INFORMATION:
; APPLICANT: CARLOW, CLOTTIDE K.S.
; APPLICANT: PAGE, ANTONY
; TITLE OF INVENTION: METHOD FOR IDENTIFYING ANTI-PARASITIC
; TITLE OF INVENTION: COMPOUNDS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID G. CONLIN, DIKE, BRONSTEIN, ROBERTS &
; ADDRESS: CUSHMAN
; STREET: 130 WATER STREET
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/145,995A
; FILING DATE: 29-OCT-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: RESNICK, DAVID S.
; REGISTRATION NUMBER: 34235
; REFERENCE/DOCKET NUMBER: 43406
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 523-3400
; TELEFAX: (617) 523-6440
; TELEX: 200291 STRE UR
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 169 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-145-995A-7

Query Match 1.7%: Score 7; DB 1; Length 169;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GGOPAGR 40
|||||
DB 11 GGOPAGR 17

RESULT 34
US-08-451-747-7
; Sequence 7, Application US/08451747
; Patent No. 5821107
; GENERAL INFORMATION:
; APPLICANT: CARLOW, CLOTTIDE K.S.
; APPLICANT: PAGE, ANTONY
; TITLE OF INVENTION: METHOD FOR IDENTIFYING ANTI-PARASITIC
; TITLE OF INVENTION: COMPOUNDS

NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: GREGORY D. WILLIAMS, NEW ENGLAND BIOLABS, INC.
STREET: 32 TOZER ROAD
CITY: BEVERLY
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 01915

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/451,747
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/145,995
FILING DATE: 29-OCT-1993
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: WILLIAMS, GREGORY D.
REGISTRATION NUMBER: 30901
REFERENCE/DOCKET NUMBER: NEB-046-DIV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508) 927-5054
TELEFAX: (508) 927-1705
TELEX:

INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 169 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-451-747-7

Query Match 1.7%: Score 7; DB 2; Length 169;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GGOPAGR 40
|||||
DB 11 GGOPAGR 17

RESULT 35

US-09-134-852-7
; Sequence 7, Application US/09134852
; Patent No. 6127148
; GENERAL INFORMATION:
; APPLICANT: CARLOW, CLOTTIDE K.S.
; APPLICANT: PAGE, ANTONY
; TITLE OF INVENTION: METHOD FOR IDENTIFYING ANTI-PARASITIC
; TITLE OF INVENTION: COMPOUNDS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID G. CONLIN, DIKE, BRONSTEIN, ROBERTS &
; ADDRESS: CUSHMAN
; STREET: 130 WATER STREET
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/134,852
; FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/145,995
FILING DATE: 29-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: RESNICK, DAVID S.
REGISTRATION NUMBER: 34235
REFERENCE/DOCKET NUMBER: 43406
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 523-3400
TELEFAX: (617) 523-6440
TELEX: 200291 STR UR
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 169 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-09-134-852-7

Query Match 1.7%; Score 7; DB 3; Length 169;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 34 GGQAPGR 40
|||
Db 11 GGQAPGR 17

RESULT 36
US-09-328-352-6419
Sequence 6419, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 6419
LENGTH: 195
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-6419

Query Match 1.7%; Score 7; DB 4; Length 195;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 156 AKRSAA 162
|||
Db 86 AKRSAA 92

RESULT 37
US-09-252-991A-25479
Sequence 25479, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 25479
LENGTH: 200
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25479

Query Match 1.7%; Score 7; DB 4; Length 200;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 34 GGQAPGR 40
|||
Db 124 GGQAPGR 130

RESULT 38
US-09-252-991A-22833
Sequence 22833, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 22833
LENGTH: 262
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22833

Query Match 1.7%; Score 7; DB 4; Length 262;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 AAALAA 17
|||
Db 18 AAALAA 24

RESULT 39
US-09-107-532A-4013
Sequence 4013, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A. Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998

```

APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 4013:
SEQUENCE CHARACTERISTICS:
LENGTH: 275 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc_feature
LOCATION: (B) LOCATION 1...275
SEQUENCE DESCRIPTION: SEQ ID NO: 4013:
US-09-107-532A-4013

```

```

Query Match      1.7% Score 7; DB 4; Length 275;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 319 AVTRGAK 325
Db 70 AVTRGAK 76

```

```

RESULT 40
US-09-107-532A-5035
Sequence 5035, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSER: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 5035:
SEQUENCE CHARACTERISTICS:
LENGTH: 287 amino acids
TYPE: amino acid

```

```

TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc_feature
LOCATION: (B) LOCATION 1...287
SEQUENCE DESCRIPTION: SEQ ID NO: 5035:
US-09-107-532A-5035

```

```

Query Match      1.7% Score 7; DB 4; Length 287;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 164 VKAAQAA 170
Db 92 VKAAQAA 98

```

```

Search completed: September 8, 2003, 14:11:00
Job time : 29 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2003, 14:02:47; Search time 45 Seconds

(without alignments)
1453.230 Million cell updates/sec

Title: US-09-889-756a-2

Perfect score: 412

Sequence: 1 MAFYAFKAMRAAALAAVAL.....AAPQSGVQTASEAKTASAE 412

Scoring table: OLIGO
Gapop 60.0, Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size: 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database: A_Geneseq_19Jun03.*

1:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	412	100.0	412	21	AA190283
2	133	32.3	412	23	AAU91064
3	100	24.3	412	24	ABP78092
4	10	2.4	212	21	AA170472
5	10	2.4	222	22	AAU93326
6	9	2.2	448	22	AAU61543
7	9	2.2	19938	24	ABP76681
8	8	1.9	111	20	AA137704
9	8	1.9	171	22	AAU98889

10	8	1.9	181	23	ABU05926	M. tuberculosis an
11	8	1.9	185	21	AA134400	Gene 30 human secr
12	8	1.9	194	21	AA186471	Human gene 51-enco
13	8	1.9	306	23	AB154244	Lactococcus lactis
14	8	1.9	318	23	AB154884	Bifidobacterium lo
15	8	1.9	426	22	AB128560	Novel human diagno
16	8	1.9	442	23	AB165347	Bifidobacterium lo
17	8	1.9	478	20	AAU93215	Human methionine a
18	8	1.9	478	20	AAU93215	Human p67 homology
19	8	1.9	478	20	AAU94763	Mouse type 2 methi
20	8	1.9	478	20	AAU94764	Rat type 2 methion
21	8	1.9	478	20	AAU94765	Human type 2 methi
22	8	1.9	478	22	AB150275	EIF-2-associated p
23	8	1.9	478	22	AA128377	Human methionine a
24	8	1.9	478	24	AB176374	Human type 2 amino
25	8	1.9	478	24	AB176375	Human type 2 amino
26	8	1.9	478	24	AB176377	Human type 2 amino
27	8	1.9	478	24	AB176378	Human type 2 amino
28	8	1.9	478	24	ABU07490	Protein differenti
29	8	1.9	480	24	AB176381	Rat type 2 aminope
30	8	1.9	480	24	AB176382	Rat type 2 aminope
31	8	1.9	500	22	AA173991	Human colon cancer
32	8	1.9	545	22	AB178333	Novel human diagno
33	8	1.9	609	22	AB133539	Novel human diagno
34	8	1.9	689	22	AB130067	Novel human diagno
35	8	1.9	891	22	AAU34554	E. coli cellular p
36	8	1.9	892	22	AAU38250	Salmonella typhi c
37	8	1.9	968	22	AB161691	Drosophila melanog
38	8	1.9	1300	22	AB168075	Drosophila melanog
39	8	1.9	1882	22	AB1605084	Novel human diagno
40	8	1.9	1883	22	AAU40183	Human polypeptide
41	8	1.9	2265	23	AB151868	Human NOV103c prot
42	8	1.9	2327	23	ABU65187	Human NOV103b prot
43	8	1.9	2390	23	ABU65186	Human NOV103a prot
44	8	1.9	2517	21	AA171159	Human phosphodiester
45	8	1.9	2517	21	AB176678	Streptomyces virid
46	8	1.7	9	24	AB120046	MHC binding peptid
47	8	1.7	15	23	AAU49383	Human macro protei
48	8	1.7	15	23	AAU49383	Mycobacterium tube
49	7	1.7	15	23	AAU98653	Mycobacterium tube
50	7	1.7	17	22	AB150834	Human liver peptid
51	7	1.7	17	22	AB130802	Peptide #3453 enco
52	7	1.7	17	22	AB135979	Peptide #3485 enco
53	7	1.7	17	22	AB131389	Protein #3388 enco
54	7	1.7	17	22	AAU66781	Human brain expres
55	7	1.7	17	22	AAU69159	Human bone marrow
56	7	1.7	17	22	AAU69159	Peptide #3424 enco
57	7	1.7	17	22	AAU29482	Peptide #3519 enco
58	7	1.7	17	22	AAU29482	Peptide #3379 enco
59	7	1.7	17	22	AB138756	Human peptide enco
60	7	1.7	18	17	AA1895898	Fragment #3 of 7-a
61	7	1.7	19	22	AB135634	Collagenase cleava
62	7	1.7	37	22	AB155657	Human liver peptid
63	7	1.7	37	22	AB140379	Peptide #7885 enco
64	7	1.7	37	22	AAU61187	Human brain expres
65	7	1.7	37	22	AAU73906	Human bone marrow
66	7	1.7	37	22	AAU4085	Peptide #8122 enco
67	7	1.7	37	22	AB143794	Human peptide enco
68	7	1.7	51	22	AAU02016	Gene #40 human sec
69	7	1.7	51	22	AAU54897	Human muscleoskele
70	7	1.7	65	22	AB140411	Human muscleoskele
71	7	1.7	65	22	ABU13305	Novel human muscle
72	7	1.7	99	24	AB111893	Human novel protei
73	7	1.7	99	24	AB111893	Human novel protei
74	7	1.7	101	23	AB111319	Human novel protei
75	7	1.7	104	22	AB108601	Novel ORFX protein
76	7	1.7	105	13	AA122384	Antigen cc-26h. E
77	7	1.7	106	23	AB100215	Human ORFX protein
78	7	1.7	111	20	AAU55257	Chlamydia pneumonia
79	7	1.7	112	21	AAU55317	Human 5' EST relat
80	7	1.7	113	21	AAU41020	Human ORFX ORF784
81	7	1.7	113	23	AB133755	Human dehydrogenas
82	7	1.7	114	7	AAU60728	Synthetic elastome

83	7	1.7	114	22	AAW90521	Human immune/haema
84	7	1.7	114	22	AAO09922	Human polypeptide
85	7	1.7	116	22	AAU94642	Propionibacterium
86	7	1.7	121	21	AAW04055	Human secreted pro
87	7	1.7	122	21	AAW08770	Arabidopsis thalia
88	7	1.7	122	21	AAW43679	Arabidopsis thalia
89	7	1.7	123	22	AAU47983	Propionibacterium
90	7	1.7	123	22	AAU18822	Novel human secret
91	7	1.7	125	21	AAW08769	Arabidopsis thalia
92	7	1.7	125	21	AAW43678	Arabidopsis thalia
93	7	1.7	126	18	AAW06577	Lipo-oligosacchari
94	7	1.7	131	19	AAW42080	The amino acid seq
95	7	1.7	132	18	AAW16323	Human ARF-p19, a n
96	7	1.7	132	21	AAV79145	Human ARF-p19 cell
97	7	1.7	132	22	AAW86208	Human p14ARF tumor
98	7	1.7	132	23	AAE25911	Human ARF-p19 prot
99	7	1.7	132	23	ABP63705	Human ORF75. Homo
100	7	1.7	132	23	AAO15374	Human Arf protein.
101	7	1.7	135	24	ABP76275	Human GENSER prote
102	7	1.7	136	21	AAW44118	Arabidopsis thalia
103	7	1.7	138	22	AAW41753	Human polypeptide
104	7	1.7	144	21	AAW42090	Arabidopsis thalia
105	7	1.7	148	21	AAW01571	Human secreted pro
106	7	1.7	152	22	ABW69617	Drosophila melanog
107	7	1.7	153	21	AAW57110	Human prostate can
108	7	1.7	163	21	AAW47117	Arabidopsis thalia
109	7	1.7	166	22	AAW58218	Propionibacterium
110	7	1.7	173	21	AAW34362	Arabidopsis thalia
111	7	1.7	173	24	ABW48207	Human bladder canc
112	7	1.7	173	24	ABW56475	Lung cancer-associ
113	7	1.7	176	21	AAW45174	Gene 2 human secre
114	7	1.7	176	21	AAW45175	Human secreted pro
115	7	1.7	179	22	ABW15529	Human secreted pro
116	7	1.7	179	22	AAW41136	Human polypeptide
117	7	1.7	181	23	AAE18325	Keratin-associated
118	7	1.7	182	24	ABP78490	N. gonorrhoeae ami
119	7	1.7	197	9	AAW06683	Phosphinothricin r
120	7	1.7	200	21	AAW44117	Arabidopsis thalia
121	7	1.7	200	21	AAW47116	Arabidopsis thalia
122	7	1.7	203	21	AAW43068	Human ORF ORF2832
123	7	1.7	203	22	ABW10250	Human cDNA SEQ ID
124	7	1.7	203	22	AAW39350	Human polypeptide
125	7	1.7	203	23	ABP66837	Human polypeptide
126	7	1.7	203	23	AAW49382	Human macro protei
127	7	1.7	205	9	AAW0916	Sequence of bovine
128	7	1.7	206	22	AAW39238	Propionibacterium
129	7	1.7	210	21	AAW34361	Arabidopsis thalia
130	7	1.7	218	22	ABW62076	Drosophila melanog
131	7	1.7	225	21	AAW45121	Human secreted pro
132	7	1.7	227	22	AAW57556	Propionibacterium
133	7	1.7	233	21	AAW42089	Arabidopsis thalia
134	7	1.7	236	22	AAW92898	C glutamincum prote
135	7	1.7	252	22	AAW79289	Human protein SEQ
136	7	1.7	253	21	AAW05005	Arabidopsis thalia
137	7	1.7	255	21	AAW42088	Arabidopsis thalia
138	7	1.7	257	21	AAW47115	Arabidopsis thalia
139	7	1.7	257	23	ABP03090	Human ORF protein
140	7	1.7	259	21	AAW43535	Arabidopsis thalia
141	7	1.7	263	22	ABW16837	Novel human diapo
142	7	1.7	267	21	AAW34360	Arabidopsis thalia
143	7	1.7	269	21	AAW05087	Arabidopsis thalia
144	7	1.7	271	21	AAW77960	A. thaliana enviro
145	7	1.7	272	24	ABW78100	N. gonorrhoeae ami
146	7	1.7	275	22	AAW35249	Enterococcus faeca
147	7	1.7	275	23	AAU05677	M. tuberculosis an
148	7	1.7	279	24	ABU70807	Human adipocyte se
149	7	1.7	281	21	AAW05004	Arabidopsis thalia
150	7	1.7	281	22	AAW48050	Signal transductio
151	7	1.7	288	21	AAW3810	Arabidopsis thalia
152	7	1.7	290	21	AAW05086	Arabidopsis thalia
153	7	1.7	292	21	AAW06078	Arabidopsis thalia
154	7	1.7	295	21	AAW26372	Arabidopsis thalia
155	7	1.7	295	23	ABW48857	Listeria monocyog

Staphylococcus epi
Arabidopsis thalia
Arabidopsis trans
Arabidopsis thalia
Arabidopsis thalia
C glutamincum prote
C glutamincum prote
Arabidopsis thalia
Amino acid sequenc
Oryza sativa perox
Human cytidine dea
Neisseria meningit
Herbicidally activ
Mycobacterium tube
Mycobacterium tube
Lycopene cyclase -
Lycopene cyclase.
Ewunia herdicola
Arabidopsis thalia
N. gonorrhoeae ami
Arabidopsis thalia
C. albicans BAX-as
Mycobacterium tube
Arabidopsis thalia
Staphylococcus epi
Arabidopsis thalia
Recombinant trans
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Breast and ovarian
Arabidopsis thalia
Arabidopsis thalia
Drosophila melanog
N. takasagoensis c
FliJ0512, fliB clone
Human albumin fusi
Human secreted pro
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Drosophila melanog
Arabidopsis thalia
Human protein SEQ
Arabidopsis thalia
Arabidopsis thalia
Drosophila melanog
Arabidopsis thalia
Arabidopsis thalia
Homo sapiens secre
Human protein SEQ
Arabidopsis thalia
Arabidopsis thalia
Human PRO1917 prot
PRO1917, an inosit
Human PRO1917 poly
Human AFP protein
Human PRO1917 prot
Amino acid sequenc
Human PRO1917 poly
Novel human secret
Human secreted pro
Human angiotensin
Human PRO1917 prot
Human PRO protein,
Human PRO polypept
Human secreted/lra
Novel secreted/lra
Human secreted/lra
Human PRO1917 poly
Novel human diapo
Arabidopsis endo-1
Strawberry cell.
Human protein sequ

229	7	1.7	498	23	ABP43592	302	7	1.7	1078	23	ABB80736	Collagen type III-
230	7	1.7	500	24	ABP77032	303	7	1.7	1078	23	ABB09628	Amino acid sequenc
231	7	1.7	514	23	AAW48334	304	7	1.7	1078	23	AAE16478	Human collagen alp
232	7	1.7	515	23	ABP28186	305	7	1.7	1082	23	ABP62883	Human polypeptide
233	7	1.7	516	21	AAAG25537	306	7	1.7	1090	22	AAAB94737	Human protein sequ
234	7	1.7	516	21	AAAG43130	307	7	1.7	1098	22	ABG03347	Novel human diagn
235	7	1.7	528	23	AAU93139	308	7	1.7	1144	21	AAAB02007	Type III adenyllyl
236	7	1.7	528	23	AAU93139	309	7	1.7	1186	15	AAAR56449	Delta endotoxin fr
237	7	1.7	536	22	AAU92991	310	7	1.7	1186	18	AAAW31503	Nematode toxin 80f
238	7	1.7	536	22	ABBB61067	311	7	1.7	1186	18	AAAW21694	Bacillus thuringie
239	7	1.7	539	20	AAAY23914	312	7	1.7	1186	18	AAAM0652	Bacillus thuringie
240	7	1.7	539	23	AAU97744	313	7	1.7	1196	13	AAAR28916	Type III procollag
241	7	1.7	539	23	AAU98656	314	7	1.7	1212	21	AAAB23158	Human colorectal c
242	7	1.7	541	21	AAAG38769	315	7	1.7	1212	24	ABJ37061	Human breast cance
243	7	1.7	541	21	ABBB91617	316	7	1.7	1257	22	ABJ70682	Drosophila melanog
244	7	1.7	542	21	AAAG25536	317	7	1.7	1273	22	ABG00113	Novel human diagn
245	7	1.7	542	21	AAAG43129	318	7	1.7	1277	22	AAAB76536	Corynebacterium gl
246	7	1.7	549	22	ABG13328	319	7	1.7	1353	22	AAAW79520	Human protein SBQ
247	7	1.7	549	22	ABG25686	320	7	1.7	1367	22	AAAW78536	Human protein SBQ
248	7	1.7	555	21	AAAG38768	321	7	1.7	1466	22	ABBS0291	Collagen type III
249	7	1.7	556	22	ABP73668	322	7	1.7	1466	23	ABBS0747	Human Tumour Endot
250	7	1.7	559	22	AAW78707	323	7	1.7	1466	24	ABBA7418	Breast cancer asso
251	7	1.7	561	22	ABG27440	324	7	1.7	1466	24	ABU54454	Human tumour endot
252	7	1.7	568	22	ABG28701	325	7	1.7	1469	22	ABG15191	Novel human diagn
253	7	1.7	570	23	AAAB71312	326	7	1.7	1787	23	ABBA49791	Listeria monocycog
254	7	1.7	570	23	AAU71858	327	7	1.7	1846	22	ABBB63163	Drosophila melanog
255	7	1.7	572	19	AAW54159	328	7	1.7	2012	22	AAAB76534	Corynebacterium gl
256	7	1.7	572	22	ABBB82871	329	7	1.7	2139	22	ABBB2393	Drosophila melanog
257	7	1.7	576	22	ABG12306	330	7	1.7	2993	22	ABBB2485	C glutathion prote
258	7	1.7	579	22	ABG29084	331	7	1.7	3079	15	AAAR59926	GAP protein Itaz.
259	7	1.7	591	21	AAAB27325	332	7	1.7	4132	22	ABGG98870	S. cinnamomensis M
260	7	1.7	599	22	ABG24227	333	7	1.7	7068	22	AAE10142	Streptomyces nous
261	7	1.7	600	23	ABP65869	334	7	1.7	9477	22	AAE10144	Streptomyces nous
262	7	1.7	602	23	ABBB68917	335	7	1.7	9477	22	ABBA7131	Human Blys binding
263	7	1.7	615	21	ABBB24068	336	7	1.7	9477	8	AAAB40855	Cytotoxic Epstein-
264	7	1.7	615	21	AAAY66678	337	7	1.7	9477	9	AAAY10577	HLA Class I motif
265	7	1.7	615	22	AAAB65201	338	7	1.7	9477	9	AAAY10577	T cell epitope/MHC
266	7	1.7	615	24	ABUS9094	339	7	1.7	9477	9	ABP12579	HIV A02 super moti
267	7	1.7	615	24	ABUS9241	340	7	1.7	9477	9	ABP12580	HIV A02 super moti
268	7	1.7	615	24	ABUS9390	341	7	1.7	9477	9	ABP20808	HIV A03 motif gag
269	7	1.7	615	24	ABUS60525	342	7	1.7	9477	9	ABP20809	HIV A03 motif gag
270	7	1.7	615	24	ABUS6016	343	7	1.7	9477	9	ABG79875	MHC class I molecu
271	7	1.7	615	24	ABUS8947	344	7	1.7	9477	9	ABGG80260	MHC class I molecu
272	7	1.7	615	24	ABU13907	345	7	1.7	9477	9	ABBP6478	HLA class I molecu
273	7	1.7	615	24	ABU10862	346	7	1.7	9477	9	ABR06155	Human cancer-relat
274	7	1.7	622	23	ABG91788	347	7	1.7	9477	9	ABR06466	Human cancer-relat
275	7	1.7	631	21	AAAY59208	348	7	1.7	9477	9	ABR06675	Human cancer-relat
276	7	1.7	631	24	ABP97686	349	7	1.7	9477	9	AAAP60730	Block unit. Synth
277	7	1.7	631	24	AAO26962	350	7	1.7	9477	10	ABP15598	HIV A24 super moti
278	7	1.7	644	22	ABBB70158	351	7	1.7	9477	10	ABP15599	HIV A24 super moti
279	7	1.7	645	22	ABBB61545	352	7	1.7	9477	10	ABP18757	HIV B62 super moti
280	7	1.7	662	22	ABBS9430	353	7	1.7	9477	10	ABP18899	HIV B62 super moti
281	7	1.7	665	15	AAAR47189	354	7	1.7	9477	10	ABP20452	HIV A03 motif gag
282	7	1.7	674	22	AAAB93773	355	7	1.7	9477	10	ABP20810	HIV A03 motif gag
283	7	1.7	687	24	ABP77385	356	7	1.7	9477	10	AAAG35630	Human complementar
284	7	1.7	749	24	ABR41749	357	7	1.7	9477	10	AAO18577	Area-specific modl
289	7	1.7	749	24	ABR41749	358	7	1.7	9477	10	ABR06837	Human cancer-relat
290	7	1.7	848	17	AAAB43716	359	7	1.7	9477	11	AAAR08088	Artifreeze segment
291	7	1.7	896	17	AAAM07733	360	7	1.7	9477	11	AAAR08088	Pneumocystis carin
292	7	1.7	900	22	ABBB61260	361	7	1.7	9477	11	AAAR08088	Sic homolog 3 dom
293	7	1.7	908	22	AAAG81162	362	7	1.7	9477	11	AAAB21136	HIV A02 super moti
294	7	1.7	908	21	AAAY54055	363	7	1.7	9477	11	ABP12811	HIV A24 super moti
295	7	1.7	912	21	AAAY54057	364	7	1.7	9477	11	ABP18761	HIV A24 super moti
296	7	1.7	917	13	AAAY5334	365	7	1.7	9477	11	ABP20465	HIV B62 super moti
297	7	1.7	917	21	AAAY5073	369	7	1.7	9477	11	ABP20788	HIV A03 motif gag
298	7	1.7	970	23	ABP65559	370	7	1.7	9477	11	ABP20811	HIV A03 motif gag
299	7	1.7	1036	23	ABG91801	371	7	1.7	9477	11	ABP20812	HIV A03 motif gag
300	7	1.7	1078	16	AAAR71704	372	7	1.7	9477	12	AAAW56681	Pneumocystis carin
301	7	1.7	1078	21	AAAY96125	373	7	1.7	9477	12	ABP12909	HIV A02 super moti

375	6	1.5	12	22	ABP12910	448	6	1.5	27	17	AAR98018	Fusogenic peptide
376	6	1.5	12	22	ABP12911	449	6	1.5	27	22	AAB55313	Anti-HPV3 F1 regio
377	6	1.5	12	22	ABP12912	450	6	1.5	27	22	AAB55344	Anti-HPV3 F1 regio
378	6	1.5	12	22	ABP18746	451	6	1.5	28	16	AAR76556	HPF3 peptide deriv
379	6	1.5	12	22	ABP18900	452	6	1.5	28	16	AAR64714	HPF3 peptide deriv
380	6	1.5	13	16	AAR87410	453	6	1.5	28	21	AAV88686	Core polypeptide f
381	6	1.5	13	21	AAAB08007	454	6	1.5	28	21	AAV88687	Core polypeptide f
382	6	1.5	13	21	AAV99239	455	6	1.5	28	21	AAV88688	Core polypeptide f
383	6	1.5	13	23	AAO18843	456	6	1.5	28	22	ABBO0045	Viral DP178/107-11
384	6	1.5	13	24	ABP70245	457	6	1.5	28	22	ABBO0046	Viral DP178/107-11
385	6	1.5	14	14	AAR32963	458	6	1.5	28	22	ABBO0047	Viral DP178/107-11
386	6	1.5	15	15	AAR31652	459	6	1.5	28	22	ABBO1509	Viral core polypep
387	6	1.5	15	24	AAR31653	460	6	1.5	28	22	ABBO1510	Viral core polypep
388	6	1.5	15	24	AAR31670	461	6	1.5	28	22	ABBO1511	Viral core polypep
389	6	1.5	15	24	AAR31671	462	6	1.5	28	22	AAU12594	DP178-1like/DP107-1
390	6	1.5	15	24	AAR31875	463	6	1.5	28	22	AAU12595	DP178-1like/DP107-1
391	6	1.5	15	24	AAR32182	464	6	1.5	28	22	AAU12596	Core polypeptide T
392	6	1.5	15	24	AAR32183	465	6	1.5	28	22	AAU12597	Core polypeptide T
393	6	1.5	15	24	AAR32184	466	6	1.5	28	22	AAU12598	Core polypeptide T
394	6	1.5	15	24	AAR32277	467	6	1.5	28	22	AAU12599	Anti-HPV3 F1 regio
395	6	1.5	15	24	AAR32278	468	6	1.5	28	22	AAU12600	Anti-HPV3 F1 regio
396	6	1.5	15	24	AAR32279	469	6	1.5	28	22	AAU12601	Anti-HPV3 F1 regio
397	6	1.5	16	20	AAR64703	470	6	1.5	29	16	AAR76557	HPF3 peptide deriv
398	6	1.5	17	16	AAR64704	471	6	1.5	29	16	AAR76558	HPF3 peptide deriv
399	6	1.5	17	22	AAB55303	472	6	1.5	29	20	AAR78143	Human secreted pro
400	6	1.5	18	16	AAR64704	473	6	1.5	29	22	AAB55316	Anti-HPV3 F1 regio
401	6	1.5	18	22	AAU04221	474	6	1.5	29	22	AAB55317	Anti-HPV3 F1 regio
402	6	1.5	18	22	AAB55304	475	6	1.5	29	23	AAR76559	HPF3 peptide deriv
403	6	1.5	18	22	AAB55305	476	6	1.5	30	16	AAR76560	HPF3 peptide deriv
404	6	1.5	19	16	AAR64705	477	6	1.5	30	16	AAR76561	HPF3 peptide deriv
405	6	1.5	19	22	AAB55306	478	6	1.5	30	22	AAB55318	Novel human diagno
406	6	1.5	20	16	AAR64706	479	6	1.5	30	22	AAB55319	Novel human diagno
407	6	1.5	20	17	AAR89341	480	6	1.5	30	22	AAB55320	Anti-HPV3 F1 regio
408	6	1.5	20	18	AAW32116	481	6	1.5	30	22	AAB55321	Anti-HPV3 F1 regio
409	6	1.5	20	21	AAV49332	482	6	1.5	30	22	AAB55322	Anti-HPV3 F1 regio
410	6	1.5	20	22	AAB55306	483	6	1.5	31	16	AAR76559	HPF3 peptide deriv
411	6	1.5	21	16	AAR64707	484	6	1.5	31	16	AAR76560	HPF3 peptide deriv
412	6	1.5	21	22	AAB55307	485	6	1.5	31	22	AAB55317	Anti-HPV3 F1 regio
413	6	1.5	22	16	AAR64708	486	6	1.5	31	22	AAB55318	Anti-HPV3 F1 regio
414	6	1.5	22	20	AAV19768	487	6	1.5	31	23	AAE28517	HPF3 peptide deriv
415	6	1.5	22	21	AAB51604	488	6	1.5	32	16	AAR76560	HPF3 peptide deriv
416	6	1.5	22	22	AAB55308	489	6	1.5	32	16	AAR76561	HPF3 peptide deriv
417	6	1.5	23	16	AAR64709	490	6	1.5	32	22	AAB55319	Anti-HPV3 F1 regio
418	6	1.5	23	20	AAW92516	491	6	1.5	32	22	AAB55348	Anti-HPV3 F1 regio
419	6	1.5	23	22	AAB55309	492	6	1.5	32	24	ABP78697	N. gonorrhoeae ami
420	6	1.5	23	23	AAB55310	493	6	1.5	33	16	AAR64719	HPF3 peptide deriv
421	6	1.5	23	23	AAB55311	494	6	1.5	33	16	AAR76561	HPF3 peptide deriv
422	6	1.5	24	16	AAR64710	495	6	1.5	33	22	ABG51491	Human liver peptid
423	6	1.5	24	16	AAR64711	496	6	1.5	33	22	ABG51492	Peptide #4079 enco
424	6	1.5	24	17	AAR89353	497	6	1.5	33	22	ABG51493	Peptide #4147 enco
425	6	1.5	24	21	AAB51532	498	6	1.5	33	22	ABG51494	Protein #3974 enco
426	6	1.5	24	21	AAV49335	499	6	1.5	33	22	AAV57403	Human brain expres
427	6	1.5	24	22	AAB55310	500	6	1.5	33	22	AAV57404	Human bone marrow
428	6	1.5	24	22	AAB55311	501	6	1.5	33	22	AAV57405	Peptide #4054 enco
429	6	1.5	25	16	AAR76553	502	6	1.5	33	22	AAV57406	Peptide #4175 enco
430	6	1.5	25	16	AAR76554	503	6	1.5	33	22	AAV57407	Peptide #3960 enco
431	6	1.5	25	21	AAV96541	504	6	1.5	33	22	AAV57408	Anti-HPV3 F1 regio
432	6	1.5	25	21	AAV96542	505	6	1.5	33	22	AAV57409	Anti-HPV3 F1 regio
433	6	1.5	25	22	AAB55311	506	6	1.5	33	23	AAB55310	Human peptide enco
434	6	1.5	25	22	AAB55312	507	6	1.5	33	23	ABP2190	Streptococcus poly
435	6	1.5	25	23	ABG62524	508	6	1.5	34	16	AAR64720	HPF3 peptide deriv
436	6	1.5	25	23	ABG62525	509	6	1.5	34	16	AAR64721	HPF3 peptide deriv
437	6	1.5	25	23	ABG62526	510	6	1.5	34	17	AAV16755	Calcitonin peptide
438	6	1.5	25	23	AAE20992	511	6	1.5	34	19	AAV20686	Human neurofilamen
439	6	1.5	25	23	AAE18367	512	6	1.5	34	22	AAB55320	Anti-HPV3 F1 regio
440	6	1.5	26	16	AAR76554	513	6	1.5	34	22	AAB55321	Anti-HPV3 F1 regio
441	6	1.5	26	16	AAR76555	514	6	1.5	35	6	AAV50187	Thymosin-alpha-11
442	6	1.5	26	22	AAV99892	515	6	1.5	35	16	AAR64721	HPF3 peptide, from
443	6	1.5	26	22	AAM43691	516	6	1.5	35	20	AAV95678	Pneumocystis carin
444	6	1.5	26	22	AAB55312	517	6	1.5	35	21	AAV88684	Core polypeptide f
445	6	1.5	26	22	AAB55313	518	6	1.5	35	21	AAV88685	Core polypeptide f
446	6	1.5	27	16	AAR76555	519	6	1.5	35	21	AAV88686	Core polypeptide f
447	6	1.5	27	16	AAR64713	520	6	1.5	35	21	AAV88687	Core polypeptide f

521	6	1.5	35	21	AAV88805	Core polypeptide f	594	6	1.5	35	22	AAV77168	Core polypeptide T
522	6	1.5	35	21	AAV88806	Core polypeptide f	595	6	1.5	35	22	AAV84832	Anti-antiviral act
523	6	1.5	35	21	AAV88807	Core polypeptide f	596	6	1.5	35	22	AAV55397	Anti-HPV3 peptide
524	6	1.5	35	21	AAV88808	Core polypeptide f	597	6	1.5	36	22	ABG58245	Human liver peptid
525	6	1.5	35	21	AAV88809	Core polypeptide f	598	6	1.5	36	22	ABV42839	Peptide #10345 enc
526	6	1.5	35	21	AAV88810	Core polypeptide f	599	6	1.5	36	22	AAV63731	Human B7-1n expres
527	6	1.5	35	21	AAV88811	Core polypeptide f	600	6	1.5	36	22	AAV76545	Human bone marrow
528	6	1.5	35	21	AAV88812	Core polypeptide f	601	6	1.5	36	22	AAV36652	Peptide #10689 enc
529	6	1.5	35	21	AAV88813	Core polypeptide f	602	6	1.5	36	22	ABG45798	Human peptid enc
530	6	1.5	35	22	ABBO0043	Viral DP178/107-11	603	6	1.5	37	21	AAV58871	Arabidopsis thalia
531	6	1.5	35	22	ABBO0161	Viral DP178/107-11	604	6	1.5	37	22	AAO12442	Human polypeptide
532	6	1.5	35	22	ABBO0162	Viral DP178/107-11	605	6	1.5	37	22	AAE04214	Human gene 17 enc
533	6	1.5	35	22	ABBO0163	Viral DP178/107-11	606	6	1.5	37	22	AAE04238	Human gene 17 enc
534	6	1.5	35	22	ABBO0164	Viral DP178/107-11	607	6	1.5	37	23	ABV882209	Chimeric pPR1R ago
535	6	1.5	35	22	ABBO0165	Viral DP178/107-11	608	6	1.5	37	23	ABG64512	Human albumin fusi
536	6	1.5	35	22	ABBO0166	Viral DP178/107-11	609	6	1.5	37	23	ABG64513	Human albumin fusi
537	6	1.5	35	22	ABBO0167	Viral DP178/107-11	610	6	1.5	38	22	ABG59559	Human liver peptid
538	6	1.5	35	22	ABBO0168	Viral DP178/107-11	611	6	1.5	38	22	ABV44156	Peptide #11702 enc
539	6	1.5	35	22	ABBO0169	Viral DP178/107-11	612	6	1.5	38	22	AAV65227	Protein #9073 enc
540	6	1.5	35	22	ABBO0170	Viral DP178/107-11	613	6	1.5	38	22	AAV65227	Human brain expres
541	6	1.5	35	22	ABBO0171	Viral DP178/107-11	614	6	1.5	38	22	AAV77930	Human bone marrow
542	6	1.5	35	22	ABBO0172	Viral DP178/107-11	615	6	1.5	38	22	AAV21825	Peptide #8259 enc
543	6	1.5	35	22	ABBO1507	Viral core polypep	616	6	1.5	38	22	AAV38149	Peptide #12186 enc
544	6	1.5	35	22	ABBO1626	Viral core polypep	617	6	1.5	38	23	ABG46952	Human peptid enc
545	6	1.5	35	22	ABBO1627	Viral core polypep	618	6	1.5	39	22	ABG02667	Novel human diagn
546	6	1.5	35	22	ABBO1628	Viral core polypep	619	6	1.5	39	22	AAV22276	Human cardiovascu
547	6	1.5	35	22	ABBO1629	Viral core polypep	620	6	1.5	39	22	AAV67535	Human acid sequenc
548	6	1.5	35	22	ABBO1630	Viral core polypep	621	6	1.5	39	24	ABV99877	Breast specific po
549	6	1.5	35	22	ABBO1631	Viral core polypep	622	6	1.5	40	21	AAV09856	Arabidopsis thalia
550	6	1.5	35	22	ABBO1632	Viral core polypep	623	6	1.5	40	22	ABG48658	Human liver peptid
551	6	1.5	35	22	ABBO1633	Viral core polypep	624	6	1.5	40	22	ABV828639	Peptide #1290 enc
552	6	1.5	35	22	ABBO1634	Viral core polypep	625	6	1.5	40	22	ABV33823	Peptide #1329 enc
553	6	1.5	35	22	ABBO1635	Viral core polypep	626	6	1.5	40	22	ABV19265	Protein #1264 enc
554	6	1.5	35	22	ABBO1636	Viral core polypep	627	6	1.5	40	22	AAV54589	Human brain expres
555	6	1.5	35	22	ABBO1637	Viral core polypep	628	6	1.5	40	22	AAV66995	Human bone marrow
556	6	1.5	35	22	AAU12552	DP178-11ike/DP107-1	629	6	1.5	40	22	AAV14855	Peptide #1289 enc
557	6	1.5	35	22	AAU12710	DP178-11ike/DP107-1	630	6	1.5	40	22	AAV27282	Peptide #1319 enc
558	6	1.5	35	22	AAU12711	DP178-11ike/DP107-1	631	6	1.5	40	22	AAV02580	Peptide #1262 enc
559	6	1.5	35	22	AAU12712	DP178-11ike/DP107-1	632	6	1.5	40	22	ABG36652	Human peptid enc
560	6	1.5	35	22	AAU12713	DP178-11ike/DP107-1	633	6	1.5	41	22	ABV57857	Human liver peptid
561	6	1.5	35	22	AAU12714	DP178-11ike/DP107-1	634	6	1.5	41	22	ABV42435	Peptide #9941 enc
562	6	1.5	35	22	AAU12715	DP178-11ike/DP107-1	635	6	1.5	41	22	ABV25885	Protein #7884 enc
563	6	1.5	35	22	AAU12716	DP178-11ike/DP107-1	636	6	1.5	41	22	AAV61321	Human brain expres
564	6	1.5	35	22	AAU12717	DP178-11ike/DP107-1	637	6	1.5	41	22	AAV76136	Human bone marrow
565	6	1.5	35	22	AAU12718	DP178-11ike/DP107-1	638	6	1.5	41	22	AAV36234	Peptide #10283 enc
566	6	1.5	35	22	AAU12719	DP178-11ike/DP107-1	639	6	1.5	41	23	ABG45502	Human peptid enc
567	6	1.5	35	22	AAU12720	DP178-11ike/DP107-1	640	6	1.5	42	20	AAV00315	Human secreted pro
568	6	1.5	35	22	AAU12721	DP178-11ike/DP107-1	641	6	1.5	43	22	AAO04607	Human polypeptide
569	6	1.5	35	22	AAU14037	DP107-1like peptid	642	6	1.5	43	22	AAO06106	Human polypeptide
570	6	1.5	35	22	AAU14038	DP107-1like peptid	643	6	1.5	43	22	AAV62281	Human gene 20-enco
571	6	1.5	35	22	AAU14039	DP107-1like peptid	644	6	1.5	44	21	AAV95964	Construct used in
572	6	1.5	35	22	AAU14040	DP107-1like peptid	645	6	1.5	44	21	AAV09855	Arabidopsis thalia
573	6	1.5	35	22	AAU14041	DP107-1like peptid	646	6	1.5	44	21	AAV83064	F-box motif of FBP
574	6	1.5	35	22	AAU14042	DP107-1like peptid	647	6	1.5	44	21	AAV24848	Human F-box motif
575	6	1.5	35	22	AAU14043	DP107-1like peptid	648	6	1.5	44	21	AAV51814	Gene 29 human secr
576	6	1.5	35	22	AAU14044	DP107-1like peptid	649	6	1.5	46	21	ABV51926	Novel human diagn
577	6	1.5	35	22	AAU14045	DP107-1like peptid	650	6	1.5	46	22	ABG27732	Novel human diagn
578	6	1.5	35	22	AAU14046	DP107-1like peptid	651	6	1.5	46	22	AAO04462	Human polypeptide
579	6	1.5	35	22	AAU14047	DP107-1like peptid	652	6	1.5	47	23	ABV25481	Streptococcus poly
580	6	1.5	35	22	AAU14048	DP107-1like peptid	653	6	1.5	48	22	ABG24708	Novel human diagn
581	6	1.5	35	22	AAV82291	Virus related pept	654	6	1.5	49	22	AAV58330	Human brain expres
582	6	1.5	35	22	AAV77039	Core polypeptide T	655	6	1.5	49	22	AAV70806	Human bone marrow
583	6	1.5	35	22	AAV77157	Core polypeptide T	656	6	1.5	49	22	AAO04481	Human polypeptide
584	6	1.5	35	22	AAV77158	Core polypeptide T	657	6	1.5	49	22	AAV18647	Peptide #5081 enc
585	6	1.5	35	22	AAV77159	Core polypeptide T	658	6	1.5	50	22	ABV16627	Human nervous syst
586	6	1.5	35	22	AAV77160	Core polypeptide T	659	6	1.5	51	21	AAV34355	Human secreted pro
587	6	1.5	35	22	AAV77161	Core polypeptide T	660	6	1.5	51	22	AAV007406	Human polypeptide
588	6	1.5	35	22	AAV77162	Core polypeptide T	661	6	1.5	52	22	ABG53747	Human liver peptid
589	6	1.5	35	22	AAV77163	Core polypeptide T	662	6	1.5	52	22	AAV56076	Propionibacterium
590	6	1.5	35	22	AAV77164	Core polypeptide T	663	6	1.5	52	22	ABV38853	Peptide #6359 enc
591	6	1.5	35	22	AAV77165	Core polypeptide T	664	6	1.5	52	22	AAV59498	Human brain expres
592	6	1.5	35	22	AAV77166	Core polypeptide T	665	6	1.5	52	22	AAV72063	Human bone marrow
593	6	1.5	35	22	AAV77167	Core polypeptide T	666	6	1.5	52	22	AAV85536	Human immune/haema

667	6	1.5	52	22	AAW32328	Peptide #6365 enco	740	6	1.5	67	19	AAW61094	Rat endogenous A6L
668	6	1.5	52	23	ABG41878	Human ORPX protein	741	6	1.5	67	22	ABW03926	Human musculoskele
669	6	1.5	52	23	ABP03591	Human ORPX protein	742	6	1.5	67	23	AAE20986	Human DNA polymera
670	6	1.5	53	21	AAW26631	Human secreted thalia	743	6	1.5	67	24	ABU13220	Novel human muscul
671	6	1.5	53	21	AAW01214	Human secreted pro	744	6	1.5	67	24	ABP80655	N. gonorrhoeae ami
672	6	1.5	53	22	AAU49365	Proionibacterium	745	6	1.5	68	19	AAW79441	Staphylococcus aur
673	6	1.5	53	22	ABW10593	Human pancreatic c	746	6	1.5	68	19	AAW61093	Human endogenous A
674	6	1.5	53	22	AAW92346	Human digestive sy	747	6	1.5	68	21	AAW59221	Human polypeptide
675	6	1.5	53	23	ABP02094	Human ORPX protein	748	6	1.5	68	23	AAE20987	Human DNA polymera
676	6	1.5	54	22	ABG52592	Human liver peptid	749	6	1.5	68	23	ABR48066	Human secreted pro
677	6	1.5	54	22	AAU50583	Proionibacterium	750	6	1.5	68	24	ABR48066	Human secreted pro
678	6	1.5	54	22	ABW37777	Peptide #5283 enco	751	6	1.5	69	8	AAW70041	Secretory signal s
679	6	1.5	54	22	ABW37777	Protein #5060 enco	752	6	1.5	69	21	AAW65529	Arabidopsis thalia
680	6	1.5	54	22	AAU1671	Novel human secret	753	6	1.5	69	21	AAW59220	Human polypeptide
681	6	1.5	54	22	AAW58391	Human brain expres	754	6	1.5	69	22	AAW12636	Human polypeptide
682	6	1.5	54	22	AAW70876	Human bone marrow	755	6	1.5	69	22	AAW2640	Human protein sequ
683	6	1.5	54	22	AAW07247	Human polypeptide	756	6	1.5	69	23	ABP34734	Human ORP3707 prot
684	6	1.5	54	22	AAW12262	Peptide #5146 enco	757	6	1.5	70	14	AAW39693	Mutant peptide der
685	6	1.5	54	22	AAW18712	Peptide #5205 enco	758	6	1.5	70	21	AAW18679	Mutant peptide der
686	6	1.5	54	22	AAW31168	Human peptide enco	759	6	1.5	70	21	AAW18690	Arabidopsis thalia
687	6	1.5	54	23	ABG40671	Human haematologic	760	6	1.5	70	21	AAW36198	Novel human diagn
688	6	1.5	55	22	AAW81169	Human haematologic	761	6	1.5	70	21	AAW36198	Novel human diagn
689	6	1.5	55	22	AAW11362	Human nervous syst	762	6	1.5	70	22	ABG25188	Human haematologic
690	6	1.5	55	22	AAW67538	Amino acid sequenc	763	6	1.5	70	22	AAW62001	DP107-like peptide
691	6	1.5	55	21	AAW09854	Arabidopsis thalia	764	6	1.5	70	22	AAW14036	HIV antiviral act
692	6	1.5	56	22	ABG53716	Human liver peptid	765	6	1.5	70	22	AAW54814	Human ovary specif
693	6	1.5	56	22	AAW48344	Proionibacterium	766	6	1.5	70	23	ABW03717	Human secreted pro
694	6	1.5	56	22	ABW38824	Peptide #6330 enco	767	6	1.5	71	21	ABW1797	Human testicular a
695	6	1.5	56	22	ABW17987	Human nervous syst	768	6	1.5	71	21	ABW6473	Human secreted pro
696	6	1.5	56	22	ABW3845	Protein #5844 enco	769	6	1.5	71	22	AAW42086	Novel human secret
697	6	1.5	56	22	AAW59468	Human brain expres	770	6	1.5	71	22	AAW33214	Novel human secret
698	6	1.5	56	22	AAW72032	Human bone marrow	771	6	1.5	71	22	AAW33214	Human reproductive
699	6	1.5	56	22	AAW82758	Human immune/haema	772	6	1.5	71	22	AAW33302	Human DTHP enzyme
700	6	1.5	56	22	AAW32298	Peptide #6335 enco	773	6	1.5	71	22	AAW59442	Proionibacterium
701	6	1.5	56	23	ABG41847	Human peptide enco	774	6	1.5	71	24	ABR41182	Novel human diagn
702	6	1.5	56	23	ABP04503	Human ORPX protein	775	6	1.5	72	22	AAW65537	Novel human diagn
703	6	1.5	57	22	AAU20698	Human novel foetal	776	6	1.5	72	22	AAW65537	Novel human diagn
704	6	1.5	57	23	AAE20982	Human DNA polymera	777	6	1.5	72	22	ABG13874	Human nervous syst
705	6	1.5	57	23	ABP26366	Streptococcus poly	778	6	1.5	72	22	ABG13874	Human nervous syst
706	6	1.5	58	22	AAU36171	Klebsiella pneumon	779	6	1.5	72	22	ABW14726	Human colon cancer
707	6	1.5	58	22	AAO00928	Human polypeptide	780	6	1.5	72	22	AAW75353	Human secreted pro
708	6	1.5	58	22	AAO12941	Human polypeptide	781	6	1.5	72	23	ABP62133	Arabidopsis thalia
709	6	1.5	58	22	AAW67534	Amino acid sequenc	782	6	1.5	73	21	AAW51998	Human secreted pro
710	6	1.5	58	22	AAE20983	Human DNA polymera	783	6	1.5	73	21	AAW34411	C glutamicum prote
711	6	1.5	59	19	AAW63840	Plasmid pRM100 Ic	784	6	1.5	73	22	AAW91348	Arabidopsis thalia
712	6	1.5	59	19	AAW63767	H. pylori strain J	785	6	1.5	74	21	AAW61539	Proionibacterium
713	6	1.5	59	20	AAW26093	Helicobacter pylor	786	6	1.5	74	22	AAW55521	Human nervous syst
714	6	1.5	59	20	AAW26099	Helicobacter pylor	787	6	1.5	74	22	ABW17492	Novel human secret
715	6	1.5	59	20	AAW26100	Helicobacter pylor	788	6	1.5	74	22	AAW32100	Human polypeptide
716	6	1.5	59	20	AAW26101	Helicobacter pylor	789	6	1.5	74	23	AAW05598	E. coli strain O15
717	6	1.5	59	22	ABW10556	Human pancreatic c	790	6	1.5	74	23	ABW82187	Toxoplasma gondii
718	6	1.5	59	22	AAW92092	Human digestive sy	791	6	1.5	75	12	AAW12347	Human secreted pro
719	6	1.5	60	9	AAW80634	Sequence encoded	792	6	1.5	75	20	AAW27673	Human DTHP recept
720	6	1.5	60	22	AAW65254	Proionibacterium	793	6	1.5	76	22	ABW14829	Human nervous syst
721	6	1.5	60	22	AAW66264	Proionibacterium	794	6	1.5	76	22	ABW14829	Human nervous syst
722	6	1.5	61	20	AAW88545	Secreted protein e	795	6	1.5	76	22	ABW14829	Human nervous syst
723	6	1.5	61	22	AAW52600	Proionibacterium	796	6	1.5	77	20	AAW87978	Human secreted pro
724	6	1.5	61	22	ABW50312	Human secreted pro	797	6	1.5	77	21	AAW00099	Novel human secret
725	6	1.5	61	22	AAO12430	Human polypeptide	798	6	1.5	77	22	AAW31667	Human rhinovirus t
726	6	1.5	61	22	ABP07173	Human ORPX protein	799	6	1.5	77	23	AAE15871	Human liver peptid
727	6	1.5	62	21	AAW26630	Arabidopsis thalia	800	6	1.5	78	22	AAW64043	Proionibacterium
728	6	1.5	62	21	AAW26630	Human polypeptide	801	6	1.5	78	22	AAW64043	Peptide #4457 enco
729	6	1.5	63	22	AAO08532	Human polypeptide	802	6	1.5	78	22	ABW31806	Peptide #4543 enco
730	6	1.5	63	23	ABP07828	Human ORPX protein	803	6	1.5	78	22	ABW37037	Protein #4351 enco
731	6	1.5	64	22	AAO08423	Simian (sal1) rota	804	6	1.5	78	22	ABW23352	Human brain expres
732	6	1.5	64	23	ABP34547	Human ORF520 prot	805	6	1.5	78	22	AAW70179	Human bone marrow
733	6	1.5	65	20	AAW13119	Human acid sequenc	806	6	1.5	78	22	AAW18002	Peptide #4436 enco
734	6	1.5	65	22	AAW67537	Human acid sequenc	807	6	1.5	78	22	AAW18002	Peptide #4550 enco
735	6	1.5	65	23	ABP35016	Human ORP3989 prot	808	6	1.5	78	22	AAW05643	Peptide #4325 enco
736	6	1.5	66	18	AAW27858	Staphylococcus aur	809	6	1.5	78	22	ABW39813	Human peptide enco
737	6	1.5	66	22	AAW90086	C glutamicum prote	810	6	1.5	79	20	AAW35258	Chlamydia pneumonia
738	6	1.5	66	22	AAW67539	Amino acid sequenc	811	6	1.5	79	20	AAW35258	Human liver peptid
739	6	1.5	66	23	ABP39721	Staphylococcus epi	812	6	1.5	79	22	ABW55647	

813	6	1.5	79	22	AAU62945	Progonibacterium
814	6	1.5	79	22	ABG24008	Novel human diagno
815	6	1.5	79	22	ABBI1531	Human nervous syst
816	6	1.5	79	22	AAU73896	Human bone marrow
817	6	1.5	79	22	AAU34075	Peptide #8112 enco
818	6	1.5	79	22	ABG43784	Human peptide enco
819	6	1.5	79	23	ABP33869	Human ORF2842 prot
820	6	1.5	80	20	ABP05241	Human ORFX protein
821	6	1.5	80	20	AAV01439	Secreted protein e
822	6	1.5	80	21	AAV87242	Human signal pepti
823	6	1.5	80	23	ABP62911	Human polypeptide
824	6	1.5	81	20	AAV60236	Human endometrium
825	6	1.5	81	22	ABBS59810	Drosophila melanog
826	6	1.5	81	22	ABB67211	Drosophila melanog
827	6	1.5	81	22	AAU28107	Novel human secret
828	6	1.5	81	23	AAO28800	Protein of drug me
829	6	1.5	81	23	ABP07049	Human ORFX protein
830	6	1.5	82	21	AAU26629	Arabidopsis thalia
831	6	1.5	82	22	AAU66762	Proionbacterium
832	6	1.5	82	22	AAU32178	Novel human secret
833	6	1.5	82	22	AAO06615	Human polypeptide
834	6	1.5	83	21	AAU24280	Arabidopsis thalia
835	6	1.5	83	21	AAU03703	Human secreted pro
836	6	1.5	83	22	ABG05602	Novel human diagno
837	6	1.5	83	22	AAO12440	Human polypeptide
838	6	1.5	83	22	AAO12487	Human polypeptide
839	6	1.5	83	23	ABP31783	Human ORF756 prote
840	6	1.5	84	23	ABP11274	Human ORFX protein
841	6	1.5	84	20	AAV04950	Mycobacterium spec
842	6	1.5	84	21	AAU13328	Human secreted pro
843	6	1.5	84	21	AAU18681	Zea mays protein f
844	6	1.5	84	21	AAU25673	Arabidopsis thalia
845	6	1.5	84	21	AAU54842	Arabidopsis thalia
846	6	1.5	84	22	AAO11259	Human polypeptide
847	6	1.5	85	21	AAU51997	Arabidopsis thalia
848	6	1.5	85	22	ABG20494	Novel human diagno
849	6	1.5	85	22	AAU84189	Human immune/hema
850	6	1.5	85	22	AAU46468	B. subtilis surfac
851	6	1.5	85	22	AAU46471	B. subtilis surfac
852	6	1.5	85	23	ABP40555	Staphylococcus epi
853	6	1.5	85	23	ABP02976	Human ORFX protein
854	6	1.5	85	24	ABP71850	Human endothelial
855	6	1.5	86	21	AAU32570	Eucalyptus grandis
856	6	1.5	86	22	ABG01722	Novel human diagno
857	6	1.5	86	22	ABG29632	Novel human diagno
858	6	1.5	86	22	AAU82819	Human immune/hema
859	6	1.5	86	23	ABU79205	Human prostate spe
860	6	1.5	86	23	ABP05717	Human ORFX protein
861	6	1.5	86	23	AAO19970	Human G-protein su
862	6	1.5	87	22	AAO08505	Human polypeptide
863	6	1.5	87	22	AAU89759	C. glutamicum prote
864	6	1.5	87	22	AAU82188	MMV p24 protein c
865	6	1.5	87	22	AAU33256	Mason-Pfizer monke
866	6	1.5	87	22	AAU75309	Human secreted pro
867	6	1.5	87	22	AAU75339	Human secreted pro
868	6	1.5	87	23	ABG60107	Human DTHP polype
869	6	1.5	88	22	AAU49877	Proionbacterium
870	6	1.5	88	22	AAU94352	Human reproductive
871	6	1.5	88	22	AAU42249	Human breast or ov
872	6	1.5	89	22	AAU59633	Proionbacterium
873	6	1.5	89	22	AAU28295	Novel human secret
874	6	1.5	89	23	ABP09354	Human ORFX protein
875	6	1.5	90	21	AAU13884	Arabidopsis thalia
876	6	1.5	90	21	AAU51275	Arabidopsis thalia
877	6	1.5	90	21	AAU01931	Human secreted pro
878	6	1.5	90	21	AAU03382	Human secreted pro
879	6	1.5	90	22	AAU65712	Human polymeric im
880	6	1.5	90	22	AAU22478	Novel human colon
881	6	1.5	90	22	AAU91946	Human digestive sy
882	6	1.5	90	22	AAO01147	Human polypeptide
883	6	1.5	90	23	AAU24325	Human lung-specifi
884	6	1.5	90	23	AAU87946	Human PDZ1 domain
885	6	1.5	91	22	AAU63056	Proionbacterium
886	6	1.5	91	22	AAU60860	Lipid biosynthesis
887	6	1.5	91	22	AAU63306	P. patens lipid met
888	6	1.5	91	23	AAU87866	Human PDZ domain #
889	6	1.5	93	22	AAU32013	Novel human secret
890	6	1.5	93	22	AAU82999	Human immune/hema
891	6	1.5	94	20	AAU26096	Helicobacter pylori
892	6	1.5	94	20	AAU26103	Helicobacter pylori
893	6	1.5	94	21	AAU65579	Human prostate can
894	6	1.5	94	21	AAU33225	Eucalyptus grandis
895	6	1.5	94	21	AAU38888	Arabidopsis thalia
896	6	1.5	94	21	AAU59219	Arabidopsis thalia
897	6	1.5	94	22	ABG11744	Novel human diagno
898	6	1.5	94	22	ABU11557	Human secreted pro
899	6	1.5	94	22	AAU22635	Novel human colon
900	6	1.5	94	22	AAU11880	Novel human secret
901	6	1.5	94	22	AAU32086	Novel human secret
902	6	1.5	94	22	AAU89692	Human immune/hema
903	6	1.5	94	22	AAU92641	Human digestive sy
904	6	1.5	94	24	ABP55311	Human polyclonog
905	6	1.5	95	21	AAU34447	Human cancer assoc
906	6	1.5	95	21	AAU65528	Arabidopsis thalia
907	6	1.5	95	21	AAU61597	Arabidopsis thalia
908	6	1.5	95	21	AAU31996	Arabidopsis thalia
909	6	1.5	95	21	AAU9636	Novel human secret
910	6	1.5	95	23	ABP32423	Human ORF1396 prot
911	6	1.5	96	18	AAU25793	Leishmania major S
912	6	1.5	96	18	AAU25794	Leishmania major S
913	6	1.5	96	22	AAU61450	Proionbacterium
914	6	1.5	96	22	AAU32014	Novel human secret
915	6	1.5	96	22	AAU94348	Human reproductive
916	6	1.5	96	22	AAU42246	Human breast or ov
917	6	1.5	96	23	ABP07289	Human ORFX protein
918	6	1.5	97	22	ABG01191	Novel human diagno
919	6	1.5	97	22	AAU31288	Novel human secret
920	6	1.5	97	22	AAU85261	Human immune/hema
921	6	1.5	97	23	ABP58508	Human topoisomeras
922	6	1.5	98	23	AAO03155	Human polypeptide
923	6	1.5	98	23	ABP53680	Human signal pepti
924	6	1.5	99	22	AAU67536	Amino acid sequenc
925	6	1.5	100	22	AAU31707	Novel human secret
926	6	1.5	100	24	ABP79361	N. gonorrhoeae ami
927	6	1.5	101	22	ABG51299	Human liver peptid
928	6	1.5	101	22	ABU69053	Drosophila melanog
929	6	1.5	101	22	ABU36459	Peptide #3965 enco
930	6	1.5	101	22	ABU1802	Protein #3801 enco
931	6	1.5	101	22	AAU78848	Human natriuretic
932	6	1.5	101	22	AAU69621	Human bone marrow
933	6	1.5	101	22	AAU29957	Peptide #3994 enco
934	6	1.5	101	22	AAU46998	P. patens S-adenos
935	6	1.5	101	23	ABG39242	Human polypeptide
936	6	1.5	101	23	ABP32361	Human ORF1334 prot
937	6	1.5	101	23	ABP26410	Streptococcus poly
938	6	1.5	101	23	ABP28552	Streptococcus poly
939	6	1.5	101	23	ABU55442	Lactococcus lactis
940	6	1.5	101	24	ABP75524	Human secretory po
941	6	1.5	101	24	ABP75509	Human secretory po
942	6	1.5	102	20	AAU04951	Mycobacterium spec
943	6	1.5	102	22	ABG12056	Novel human diagno
944	6	1.5	102	22	AAU06554	Human polypeptide
945	6	1.5	102	22	AAU83368	Human protein sequ
946	6	1.5	103	22	ABG11930	Novel human diagno
947	6	1.5	103	22	AAU11518	Novel human secret
948	6	1.5	103	22	AAU86997	Human immune/hema
949	6	1.5	103	22	AAU24355	Human EST encoded
950	6	1.5	103	22	AAU40644	Human polypeptide
951	6	1.5	103	23	AAU28648	Human K-beta2 pro
952	6	1.5	103	23	ABU82186	E. coli strain K12
953	6	1.5	104	21	AAU40916	Human ORFX ORF680
954	6	1.5	104	21	AAU28787	Arabidopsis thalia
955	6	1.5	104	22	ABG27333	Novel human diagno
956	6	1.5	104	23	ABG69820	Human secretory pr
957	6	1.5	104	23	ABP06516	Human ORFX protein
958	6	1.5	104	23	ABP08254	Human ORFX protein

959	6	1.5	105	18	AAW25792	Leishmania major S
960	6	1.5	105	22	AAU64044	Procionibacterium
961	6	1.5	106	22	AAU44465	Procionibacterium
962	6	1.5	107	21	AAU10493	N. meningitidis BA
963	6	1.5	107	21	AAV74672	Neisseria meningit
964	6	1.5	107	21	AAV74673	Neisseria meningit
965	6	1.5	107	22	AAU63357	Human breast cancer
966	6	1.5	107	22	AAU63357	Human breast cancer
967	6	1.5	107	22	ABB89818	Human polypeptide
968	6	1.5	107	24	ABP99939	Breast specific po
969	6	1.5	108	21	AAU14592	Arabidopsis thalia
970	6	1.5	108	21	AAU00317	Human secreted pro
971	6	1.5	108	22	ABG15477	Novel human diagno
972	6	1.5	108	22	AAU85779	Human immune/haema
973	6	1.5	109	23	ABP66641	Human breast speci
974	6	1.5	109	15	AAU51504	Plin protein Vari
975	6	1.5	109	22	AAU92415	C glutamic prote
976	6	1.5	109	22	AAU76631	Corynebacterium gl
977	6	1.5	109	23	ABP10815	Human ORF protein
978	6	1.5	110	22	AAU03790	Human polypeptide
979	6	1.5	110	23	ABB82188	S. typhimurium cyt
980	6	1.5	111	22	ABG14220	Novel human diagno
981	6	1.5	111	24	ABP80244	N. gonorrhoeae ami
982	6	1.5	112	21	AAU12364	Fragment of human
983	6	1.5	112	22	AAU32786	Novel human secret
984	6	1.5	113	22	AAU32788	Novel human secret
985	6	1.5	114	22	AAU00257	Human polypeptide
986	6	1.5	114	24	ABP79769	N. gonorrhoeae ami
987	6	1.5	115	18	AAU20848	H. pylori cytoplasm
988	6	1.5	115	23	ABG70160	Human prey protein
989	6	1.5	116	21	AAU42417	Human ORF ORF2181
990	6	1.5	116	22	AAU004138	Human polypeptide
991	6	1.5	116	23	ABP99941	Human transcrip
992	6	1.5	116	23	ABP08252	Human ORF protein
993	6	1.5	117	22	AAU99646	Human excretory re
994	6	1.5	117	22	AAU42461	Human kidney relat
995	6	1.5	118	20	AAU16956	Human fascicular a
996	6	1.5	118	20	AAU12314	Human 5' EST seque
997	6	1.5	118	22	ABP61583	Procionibacterium
998	6	1.5	118	23	ABP57818	Human synaptic ves
999	6	1.5	119	22	AAU87677	Novel central nerv
1000	6	1.5	119	22	AAU68016	Procionibacterium
	6	1.5	119	22	AAU04440	Human polypeptide

ALIGNMENTS

RESULT 1
AAU90283
ID AAY90283 standard; Protein; 412 AA.

XX AAY90283;

DT 24-OCT-2000 (first entry)

XX N. meningitidis BASB05 protein sequence.

XX BASB05; diagnosis; microbial infection; invasive bacterial disease;
KW Neisseria meningitidis infection; upper respiratory tract infection;
KW bacteraemia; meningitis; therapy.

XX Neisseria meningitidis.

XX WO200043517-A1.

XX 27-JUL-2000.

PF 19-JAN-2000; 2000WO-EP00425.

XX 22-JAN-1999; 99GB-0001462.

PR 29-JAN-1999; 99GB-0002069.

PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX	Thomard J;	100.0%; Score 412; DB 21; Length 412;
PI	Best Local Similarity 100.0%; Pred. No. 0;	
XX	Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
DR	WPI: 2000-476199/41.	
DR	N-PSDB; AAU37645.	
XX	Isolated BASB05 polypeptides, polynucleotides, and antibodies, the	
PT	polypeptides and polynucleotides are useful as vaccines for treating	
PT	infection and diagnosing a microbial infection such as a Neisseria meningitidis	
XX	infection	
PS	Claim 3, Page 63; 82pp; English.	
XX	This sequence represents the Neisseria meningitidis BASB05 polypeptide	
CC	of the invention. The BASB05 polypeptides and polynucleotides are useful	
CC	for diagnosing and treating microbial infections such as a Neisseria	
CC	meningitidis infection. They can also be used to treat any disease caused	
CC	by or related to infection by a bacterium, including upper respiratory	
CC	tract infection, invasive bacterial diseases (such as bacteraemia) and	
CC	meningitis.	
XX	Sequence 412 AA:	
SO	Query Match 100.0%; Score 412; DB 21; Length 412;	
	Best Local Similarity 100.0%; Pred. No. 0;	
	Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	

QY	1	MAFYAFKMRRAAALAAVALVLSGCKGDAAGSGPAGREAPAVGVVTHPQTVALT	60
DB	1	MAFYAFKMRRAAALAAVALVLSGCKGDAAGSGPAGREAPAVGVVTHPQTVALT	60
QY	61	VELPRLSLRTADYRAVGGIIQKRLFOESYVAGQPLVQIDSTYEAUMLSARAQLA	120
DB	61	VELPRLSLRTADYRAVGGIIQKRLFOESYVAGQPLVQIDSTYEAUMLSARAQLA	120
QY	121	TAQATLAAADADLARYKPLVAEAVSROEYPAATYAKRSAGVAAQAQAIKSAGINLR	180
DB	121	TAQATLAAADADLARYKPLVAEAVSROEYPAATYAKRSAGVAAQAQAIKSAGINLR	180
QY	181	SRITPPIGFTGQSVSEGTLLNMGDTVLTATITQTNMYNVNQSASEWKLRLROIAG	240
DB	181	SRITPPIGFTGQSVSEGTLLNMGDTVLTATITQTNMYNVNQSASEWKLRLROIAG	240
QY	241	KLADGVIAVGIRKPDGTVYPEKGRLLFADPVNVESTGQITLRAAVNDONILMPGLYV	300
DB	241	KLADGVIAVGIRKPDGTVYPEKGRLLFADPVNVESTGQITLRAAVNDONILMPGLYV	300
QY	301	RYLMDQVAVDNAFVVPQAVTRGAKDTVMYNAQCGMEPREVTVAQOQGTNMTVTSGLKD	360
DB	301	RYLMDQVAVDNAFVVPQAVTRGAKDTVMYNAQCGMEPREVTVAQOQGTNMTVTSGLKD	360
QY	361	GDKVVVEGISTAGITGAKKVPKEMWASSENQAAPOSGVOTASEKTASEAE	412
DB	361	GDKVVVEGISTAGITGAKKVPKEMWASSENQAAPOSGVOTASEKTASEAE	412

RESULT 2

ID AAU91064 standard; Protein; 412 AA.

XX AAU91064;

DT 05-JUN-2002 (first entry)

XX Neisseria cell surface polypeptide #2.

XX Cell surface protein; antibacterial; antimicrobial.

XX Neisseria meningitidis.

XX WO200216612-A2.

PD 28-FEB-2002.

```

XX 21-AUG-2001; 2001WO-GB03759.
XX
XX 24-AUG-2000; 2000GB-0020952.
XX
XX (MIGR-) MICROSCIENCE LTD.
XX
XX Lane JD, Hughes MJG, Santangelo JD;
XX
XX WPI: 2002-280941/32.
XX N-PSDB; ABEK54080.
XX
XX Novel peptide encoded by Neisseria meningitidis, useful for manufacture
XX of medicament for treatment or prevention of condition associated with
XX infection by Neisseria or Gram-negative bacteria -
XX
XX Claim 5, Page 19-20; 79pp; English.
XX
XX The invention relates to polypeptides located on the cell surface of
XX Neisseria meningitidis, and the polynucleotides encoding them. The
XX sequences of the invention are useful for therapeutic or diagnostic use,
XX in the manufacture of a medicament for use in treatment or prevention of
XX a condition associated with infection by Neisseria or Gram-negative
XX bacteria. The sequences are also useful for screening potential
XX antimicrobial drugs or for detection of virulence. Sequences
XX AAU91063-AAU91079 represent Neisseria meningitidis polypeptides of the
XX invention.
XX
XX Sequence 412 AA;
SQ
XX
XX Query Match 32.3%; Score 133; DB 23; Length 412;
XX Best Local Similarity 100.0%; Pred. No. 2.6e-129;
XX Matches 133; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 274 VNESIGQITLRAAVPNDQNIIMPGLYVRVLMQVAVDNAFVPPQAVTGAQDTWIVNA 333
XX |||||
XX 274 VNESIGQITLRAAVPNDQNIIMPGLYVRVLMQVAVDNAFVPPQAVTGAQDTWIVNA 333
XX
XX 334 OCGMEPREVTAAQOOGTNIWITSGKDGDKVVEGISTAGTGAKKVTPKEMASSENOAA 393
XX |||||
XX 334 OCGMEPREVTAAQOOGTNIWITSGKDGDKVVEGISTAGTGAKKVTPKEMASSENOAA 393
XX
XX 394 APQSGVQTASEAK 406
XX |||||
XX 394 APQSGVQTASEAK 406
XX
XX Db
XX
XX RESULT 3
XX ABP78092
XX ID ABP78092 standard; Protein; 412 AA.
XX
XX AC ABP78092;
XX
XX DT 07-MAR-2003 (first entry)
XX
XX DE N. gonorrhoeae amino acid sequence SEQ ID 2714.
XX
XX KW Antibacterial; infection; vaccine; gene therapy.
XX
XX OS Neisseria gonorrhoeae.
XX
XX PN WO200279243-A2.
XX
XX PD 10-OCT-2002.
XX
XX PF 12-FEB-2002; 2002WO-IB02069.
XX
XX PR 12-FEB-2001; 2001GB-0003424.
XX
XX PA (CHIR-) CHIRON SPA.
XX
XX Fontana MR, Pizza M, Masignani V, Monaci E;
XX

```

```

DR WPI: 2003-058415/05.
DR N-PSDB; ABEZ39062.
XX
XX PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a
XX medicament for treating or preventing N. gonorrhoeae infection -
XX
XX PS Disclosure; Page 386; 815pp; English.
XX
XX CC The present invention relates to proteins from Neisseria gonorrhoeae.
XX Also disclosed are the nucleic acid molecules encoding the proteins and
XX antibodies that specifically bind to the proteins. The composition
XX comprising the protein, nucleic acid or antibody is useful for the
XX manufacture of a medicament for treating or preventing N. gonorrhoeae
XX infection, this may be in the form of a vaccine or gene therapy.
XX
XX CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
XX molecules of the invention.
XX
XX Sequence 412 AA;
SQ
XX
XX Query Match 24.3%; Score 100; DB 24; Length 412;
XX Best Local Similarity 100.0%; Pred. No. 5.9e-95;
XX Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 148 QEYDAAVTAKSAEKGVAQAQAIKSAGININRSRITAPISGFIQSKVSEGTLLNAGDT 207
XX |||||
XX 148 QEYDAAVTAKSAEKGVAQAQAIKSAGININRSRITAPISGFIQSKVSEGTLLNAGDT 207
XX
XX 208 TVLATIRQTNPVYVNTOSASEVMKLRQIAGKLLADG 247
XX |||||
XX 208 TVLATIRQTNPVYVNTOSASEVMKLRQIAGKLLADG 247
XX
XX Db
XX
XX RESULT 4
XX AAY70472
XX ID AAY70472 standard; Protein; 212 AA.
XX
XX AC AAY70472;
XX
XX DT 21-JUN-2000 (first entry)
XX
XX DE Human p53 target molecule, PRG6 protein.
XX
XX KW PRG6; p53 target; human; modulate; cell proliferation; immunomodulatory;
XX chromosome 19p13.2-13.1; cytostatic; gene therapy; tumour cell; inducer;
XX diagnosis; therapeutic; proliferative disease; cell cycle arrest; cancer;
XX treatment; apoptosis; knockout animal; cancer susceptibility; POU-domain;
XX homeo-domain.
XX
XX KW Homo sapiens.
XX
XX OS
XX
XX FH Key Location/Qualifiers
XX
XX FT Misc-difference 66 /note= "Encoded by TTG"
XX
XX FT Misc-difference 202 /note= "Encoded by CCA"
XX
XX PN WO200012526-A1.
XX
XX PD 09-MAR-2000.
XX
XX PF 27-AUG-1999; 99WO-US19551.
XX
XX PR 28-AUG-1998; 98US-0098251.
XX
XX PA (UYP- ) UNIV PRINCETON.
XX
XX PI Horikoshi N, Shenk T;
XX
XX WPI: 2000-246724/21.
XX N-PSDB; AAZ51674.
XX
XX New p53-inducible isolated nucleic acid molecule including open reading
XX frame encoding human homolog of Drosophila melanogaster peroxidasin,
XX

```

PT useful e.g. in detection and treatment of cancer -
XX
XX
PS Disclosure; Fig 9; 83pp; English.
XX
CC The present sequence is the protein encoded by PRG6 gene, that is
CC decreased in response to induction of p53 activity in human colon
CC cancer EBI cells. Repression of PRG6 is by a liver specific factor. It
CC is a basic protein localised in the nucleus that has POU-domain like
CC sequence and two homeo-domain like sequences. It may function as a
CC transcription factor. They are potential targets of p53 regulatory
CC activity and are useful for modulation of cellular proliferation. PRG6
CC gene is localised to human chromosome 19p13.2-13.1. The PRG target
CC molecules have cytostatic and immunomodulatory activity. PRG nucleotides,
CC proteins and antibodies are useful as diagnostic and therapeutic agents
CC for detection and treatment of cancer and other proliferative diseases.
CC The gene/cDNA may be used for gene therapy, to restore a gene function
CC downstream of p53, that cannot be activated in the p53-deficient tumour
CC cell. Antibodies can be used as inducers of cell cycle arrest and/or
CC apoptosis. The DNA sequences can be used to generate 'knockout' animals
CC as a model of cancer susceptibility.
XX
SQ Sequence 212 AA;
Query Match 2.4%; Score 10; DB 21; Length 212;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 10 RAALALAAVA 19
DB 192 RAALALAAVA 201
RESULT 5
AAG93326
ID AAG93326 standard; Protein; 222 AA.
XX
AC AAG93326;
XX
DT 13-SEP-2001 (first entry)
XX
DE Human protein HP10438.
XX
KM Human; gene therapy; tumour.
XX
OS Homo sapiens.
XX
PN WO200142302-A1.
XX
PD 14-JUN-2001.
XX
PF 06-DEC-2000; 2000WO-JP08631.
XX
PR 06-DEC-1999; 99JP-0346863.
XX
PR 06-DEC-1999; 99JP-0346864.
XX
PR 08-FEB-2000; 2000JP-0031062.
XX
PR 10-FEB-2000; 2000JP-0034090.
XX
PR 10-FEB-2000; 2000JP-0034091.
XX
PR 14-FEB-2000; 2000JP-0035829.
XX
PR 14-FEB-2000; 2000JP-0035899.
XX
PR 14-MAR-2000; 2000JP-0071161.
XX
PR 30-MAY-2000; 2000JP-0160851.
XX
PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX
PI Kato S, Eguchi C, Saeki M;
XX
DR WPI; 2001-381646/40.
XX
DR N-PSDB; AAH68611.
XX
PT Human protein originated from tumor cell line, applicable as drug,
PT reagent for studying intracellular protein networks and protein source
PT for drug screening, also encoded cDNA for gene diagnosis and gene
PT therapy -

XX
PS Claim 1; Pages 451-452; 471pp; Japanese.
XX
CC The present sequence is a human protein. The human protein, preferably
CC originated from tumor cell line, is applicable as a drug, a reagent for
CC studying intracellular protein networks and a protein source for
CC screening proteins for binding low molecular weight drugs. The human
CC protein coding sequence is useful for gene diagnosis and gene therapy,
CC expression vectors and transformant cells for detection of ligands and
CC receptors.
XX
SQ Sequence 222 AA;
Query Match 2.4%; Score 10; DB 22; Length 222;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 10 RAALALAAVA 19
DB 202 RAALALAAVA 211
RESULT 6
AAU61543
ID AAU61543 standard; Protein; 448 AA.
XX
AC AAU61543;
XX
DT 27-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #22439.
XX
KM SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KM uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KM inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KM dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US12865.
XX
PR 21-APR-2000; 2000US-199047P.
XX
PR 02-JUN-2000; 2000US-208841P.
XX
PR 07-JUL-2000; 2000US-216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAW, Perring DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI; 2001-616774/71.
XX
DR N-PSDB; AAS59618.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
PS Example 1; SEQ ID No 22738; 1069pp; English.
XX
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention

CC and determining the amount of bound protein in the sample. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC specific for P. acnes proteins. These antibodies can be used to
 CC downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA).
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

CC Sequence 448 AA;

Query Match 2.2%; Score 9; DB 22; Length 448;

Best Local Similarity 100.0%; Pred. No. 3.5;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 358 LKDGKRVV 366
 |||||

DB 147 LKDGKRVV 155

RESULT 7

ID ABP76681 standard; Protein, 19938 AA.

XX ABP76681;

DT 26-FEB-2003 (first entry)

DE Streptomyces viridochromogenes Av1 gene cluster polypeptide frame 5.

XX Avilamycin; antibacterial; virucide; protozoacide; fungicide; infection;

XX medicine; Staphylococcus aureus; biosynthetic gene cluster; enzyme.

OS Streptomyces viridochromogenes.

PN WO200268436-A1.

PD 06-SEP-2002.

PF 24-AUG-2001; 2001WO-EP09815.

XX 25-FEB-2001; 2001DE-1009166.

PA (COMB-) COMBINATURE BIOPHARM AG.

PI Weinbauer G, Muehlenweg A, Trefzer A, Bechtold A;

DR WPI; 2003-018650/01.

DR N-PSDB; AB237516.

PT New avilamycin derivatives, useful for treatment of infections, and

PT nucleic acid encoding avilamycin synthesis enzymes -

PS Example 1; Page 68-301; 319pp; German.

CC The invention relates to avilamycin derivatives (I) with antibacterial,

CC virucide, protozoacide and fungicide activity. (I) are useful for

CC treatment of infections (bacterial, viral, protozoal or fungal), in human

CC or veterinary medicine, particularly where caused by Staphylococcus

CC aureus. (I) are more hydrophilic than known avilamycins. The present

CC sequence is that of an avilamycin synthesis enzyme from the Streptomyces

CC viridochromogenes Avilamycin A biosynthetic gene cluster

CC (AB237515-AB237516).

DB 13793 AAAIAAAVA 13801
 |||||

RESULT 8

ID AAY37704 standard; Protein, 111 AA.

XX AAY37704;

DT 07-OCT-1999 (first entry)

DE Chlamydia trachomatis ribosomal polypeptide.

XX Vaccine; eye disease; conventional trachoma; nonendemic trachoma;

XX paratrachoma; inclusion conjunctivitis; genital disease; perihhepatitis;

XX nongonococcal urethritis; epididymitis; cervicitis; salpingitis;

XX Chlamydia trachomatis.

OS WO928475-A2.

PN 10-JUN-1999.

PF 27-NOV-1998; 98WO-IB01939.

XX 04-NOV-1998; 98US-0107077.

PR 28-NOV-1997; 97FR-0015041.

PR 17-DEC-1997; 97FR-0016034.

PA (GEST) GENSET.

PI Griffais R;

DR WPI; 1999-371125/31.

PT Genome sequence of Chlamydia trachomatis

PS Disclosure; Page 1317; 1755pp; English.

CC AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome

CC of Chlamydia trachomatis (see AA201425). The polypeptides can be used as

CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences

CC can also be used to control growth of the microorganism. Chlamydia

CC trachomatis is responsible for a large number of diseases, e.g. eye

CC diseases such as conventional trachoma, nonendemic trachoma,

CC paratrachoma, and inclusion conjunctivitis; genital diseases such as

CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,

CC perihhepatitis, Bartholinitis; pneumonia in breast feeding infants;

CC and venereal lymphogranulomatosis. The polypeptides of the invention

CC may be of use in treating these diseases.

SO Sequence 111 AA;

Query Match 1.9%; Score 8; DB 20; Length 111;

Best Local Similarity 100.0%; Pred. No. 9.9;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 DKVVEGI 369
 |||||

DB 32 DKVVEGI 39

RESULT 9

ID AAG98889 standard; Protein, 171 AA.

XX AAG98889;

DT 26-SEP-2001 (first entry)

DE E. coli growth and proliferation related protein sequence SEQ ID NO:359.

```

XX Escherichia coli; growth; proliferation; microbial; antimicrobial;
KW bacterial infection; microorganism.
XX
XX Escherichia coli.
OS
XX WO200134810-A2.
XX
XX 17-MAY-2001.
XX
XX 09-NOV-2000; 2000WO-US30950.
XX
XX 09-NOV-1999; 99US-0164415.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Forsyth RA, Ohlsen K, Zyskind J;
XX
XX WPI; 2001-335933/35.
XX
XX N-PSDB; AAH84560.
XX
XX Novel nucleic acids that inhibit Escherichia coli proliferation, useful
XX for screening for homologous genes and for designing expression vectors
XX
XX Claim 19; Page 435; 522pp; English.
XX
XX AAH84373 to AAH84499 represent Escherichia coli growth and proliferation
XX related DNA sequences (1). AAH84500 to AAH84670 encode the E. coli
XX growth and proliferation related proteins given in AAG99078 and AAG98830
XX to AAG99999. (1) can be used as potential targets for the generation of
XX new antimicrobial agents, and for identification of compounds which
XX interact with the gene products of (1). In addition the expression of
XX (1) and the purification of the proteins, the purified proteins can be
XX used to generate reagents and screen small molecule libraries or other
XX candidate compound libraries for compounds that can be further developed
XX to yield novel antimicrobial compounds. In addition, nucleic acid probes
XX complementary to (1) that are specific for particular species of
XX microorganisms can be used to identify particular microorganism species
XX in clinical specimens, therefore, providing a rapid and dependable
XX method by which to identify the causative agents of a bacterial
XX infection. Also, antibodies generated against proteins translated from
XX cDNA transcribed from proliferation-required sequences can also be used
XX to screen for specific microorganisms that produce such proteins in a
XX species-specific manner. AAH84371 and AAH84670 represent sequencing
XX primers used in the isolation of E. coli growth and proliferation
XX related sequence, which are used in an example from the present
XX invention.
XX
XX Sequence 171 AA;
XX
XX Query Match 1.9%; Score 8; DB 22; Length 171;
XX Best Local Similarity 100.0%; Pred. No. 15;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 201 LUNAGDTT 208
XX |||||
XX DB 61 LUNAGDTT 68
XX
XX RESULT 10
XX ABB05926
XX ID ABB05926 standard; Protein; 181 AA.
XX
XX AB05926;
XX
XX 08-APR-2003 (first entry)
XX
XX M. tuberculosis and M. leprae marker protein #577.
XX
XX Mycobacteriosis; survival; virulence; protective antigen; vaccine;
XX mycobacterial disease; tuberculosis; leprosy.
XX

```

```

OS Mycobacterium tuberculosis.
OS Mycobacterium leprae.
XX
XX WO200274903-A2.
XX
XX 26-SEP-2002.
XX
XX 22-FEB-2002; 2002WO-IB01973.
XX
XX 22-FEB-2001; 2001US-270123P.
XX
XX (INSP) INST PASTEUR.
XX
XX Cole S;
XX
XX WPI; 2002-759885/82.
XX
XX Identifying and selecting genes for survival or virulence of
XX mycobacteria by a comparative genomic analysis of the sequences of
XX Mycobacterium tuberculosis and M. leprae -
XX
XX Claim 17; Page 791; 874pp; English.
XX
XX This invention relates to a novel method for identifying essential genes
XX for survival or virulence of mycobacteria species. The method comprises
XX aligning the genomic sequence of a first mycobacterium species on a
XX genomic sequence of a second mycobacterium species and selecting a
XX polynucleotide sequence that is highly conserved in both genomes with no
XX counterparts in other bacterial genomic sequences and that corresponds
XX to an essential gene for the survival or virulence of mycobacterium
XX species. The method of the invention is useful for detecting M.
XX tuberculosis or M. leprae infection. The method reduces the number of
XX potential new targets and protective antigens for new drugs and vaccine
XX compositions to treat and prevent mycobacterial diseases, particularly
XX tuberculosis and leprosy. The present sequence represents a marker
XX protein from Mycobacterium tuberculosis and Mycobacterium leprae
XX identified using the method of the invention.
XX
XX Sequence 181 AA;
XX
XX Query Match 1.9%; Score 8; DB 23; Length 181;
XX Best Local Similarity 100.0%; Pred. No. 16;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 11 AAALAAV 18
XX |||||
XX DB 16 AAALAAV 23
XX
XX RESULT 11
XX AAB34400
XX ID AAB34400 standard; Protein; 185 AA.
XX
XX AAB34400;
XX
XX 26-JAN-2001 (first entry)
XX
XX Gene 30 human secreted protein homologous amino acid sequence #161.
XX
XX Human; secreted protein; diagnosis; neuroprotective; cyostatic;
XX cardioactive; immunomodulatory; muscular active general; vulnerrary;
XX gastrointestinal; nephrotoxic; antiinfective; gynaecological; and
XX antibacterial; gene therapy's detection; cancer; chromosome marker;
XX chromosome identification; neural disorder; immune disorder;
XX muscular disorder; reproductive disorder; gastrointestinal disorder;
XX pulmonary disorder; cardiovascular disorder; renal disorder;
XX proliferative disorder; wound healing; infectious disease; preservative;
XX food additive.
XX
XX Caenorhabditis elegans.
XX
XX WO200056883-A1.
XX

```



```

PD 28-SEP-2000.
XX
XX 16-MAR-2000; 2000WO-US06822.
PF
XX 23-MAR-1999; 99US-0126054.
PR
XX 10-DEC-1999; 99US-0169916.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM, Komatsoulis G;
XX
XX WPI; 2000-587666/55.
XX
XX Human secreted proteins and gene sequences encoding them, useful for
XX detecting, preventing, and treating disorders such as cancer,
XX neurological disorders and immune system disorders -
XX
XX Disclosure; Page 409-410; 429pp; English.
XX
XX The polynucleotide sequences given in AAC59566 to AAC59614 encode the
XX human secreted proteins given in AAB34299 to AAB34347. AAB34348 to
XX AAB34437 represent human secreted polypeptide sequences and proteins
XX homologous to them, which are given in the exemplification of the present
XX invention. Human secreted proteins have activities based on the tissues
XX and cells the genes are expressed in. Example of activities include:
XX neuroprotective; cytoskeletal; cardioactive; immunomodulatory; muscular
XX active general; vulnerary; gastrointestinal; nephrotropic;
XX antifetive; gynaecological; and antibacterial. The polynucleotides
XX can be used for the detection of various disorders such as cancer,
XX chromosome identification, as chromosome markers, and for numerous other
XX diagnostic or research purposes. The secreted proteins can be used to
XX treat disorders such as neural, immune, muscular, reproductive,
XX gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
XX disorders, wound healing, and infectious diseases. The proteins can
XX also be used as a food additive or preservative to increase or decrease
XX storage capabilities. AAC59557 to AAC59565 and AAB34238 represent
XX sequences used in the exemplification of the present invention.
XX
XX Sequence 185 AA;
XX
XX Query Match 1.9%; Score 8; DB 21; Length 185;
XX Best Local Similarity 100.0%; Pred. No. 16;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 239 ESKLAAAD 246
XX |||||
XX 141 ESKLAAAD 148
XX
XX RESULT 12
XX ID AAY86471 standard; Protein; 194 AA.
XX
XX AAY86471;
XX
XX 19-APR-2000 (first entry)
XX
XX Human gene 51-encoded protein fragment, SEQ ID NO:386.
XX
XX Human; secreted protein; cancer; tumour; developmental abnormality;
XX foetal deficiency; blood disorder; immune system disorder; inflammation;
XX autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;
XX schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder;
XX atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;
XX digestive disorder; endocrine disorder; infection; AIDS; leukaemia;
XX therapy; chromosome 19.
XX
XX Homo sapiens.
XX
XX MO9966041-A1.
XX
XX 23-DEC-1999.
XX

```

```

PF 15-JUN-1999; 99WO-US13418.
XX
XX 16-JUN-1998; 98US-0089507.
XX 16-JUN-1998; 98US-0089508.
PR
XX 16-JUN-1998; 98US-0089509.
XX 16-JUN-1998; 98US-0089510.
PR
XX 22-JUN-1998; 98US-0090112.
XX 22-JUN-1998; 98US-0090113.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Ni J, Rosen CA, Wei Y, Young PE, Florence KA,
XX Soppet DR, Brewer LA, Endress GA, Carter KC, Mucenski M, Ebner R;
XX Lafleur DW, Olsen HS, Shi Y, Moore PA, Komatsoulis G;
XX
XX WPI; 2000-106100/09.
XX
XX New isolated human genes and the secreted polypeptides they encode,
XX useful for diagnosis and treatment of e.g. cancers, neurological
XX disorders, immune diseases, inflammation or blood disorders -
XX
XX Disclosure; Page 103; 586pp; English.
XX
XX AA297019 to AA297137 represent 94 isolated human secreted protein genes.
XX AAY86215 to AAY86333 are the secreted proteins encoded by the 94 human
XX genes. This sequence represents a fragment of one of the human secreted
XX proteins. The genes and their corresponding secreted polypeptides are
XX useful for preventing, treating or ameliorating medical conditions,
XX e.g., by protein or gene therapy. Also pathological conditions can be
XX diagnosed by determining the amount of the new polypeptides in a sample
XX or by determining the presence of mutations in the new genes. Specific
XX uses are described for each of the 94 genes, based on which tissues they
XX are most highly expressed in, and include developing products for the
XX diagnosis or treatment of cancer, tumours, developmental abnormalities
XX and foetal deficiencies, blood disorders, diseases of the immune system,
XX autoimmune diseases, inflammation, allergies, Alzheimer's and cognitive
XX disorders, schizophrenia, arthritis, asthma, psoriasis, sepsis, skin
XX disorders, atherosclerosis, diabetes, cardiovascular disorders, kidney
XX disorders, digestive/endocrine disorders, infections and AIDS. The
XX polypeptides are also useful for identifying their binding partners.
XX The sequences shown in AAY86334 to AAY86585 represent fragments of the
XX secreted proteins.
XX
XX Sequence 194 AA;
XX
XX Query Match 1.9%; Score 8; DB 21; Length 194;
XX Best Local Similarity 100.0%; Pred. No. 17;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 12 AALAAAVA 19
XX |||||
XX 176 AALAAAVA 183
XX
XX RESULT 13
XX ID ABB54244 standard; Protein; 306 AA.
XX
XX ABB54244;
XX
XX 16-MAY-2002 (first entry)
XX
XX Lactococcus lactis protein yJfE.
XX
XX Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
XX
XX Lactococcus lactis IL1403.
XX
XX FR2807446-A1.
XX
XX 12-OCT-2001.
XX
XX 11-APR-2000; 2000FR-0004630.
XX

```

XX 11-APR-2000; 2000PR-0004630.
 PR (INRG) INRA INST NAT RECH AGRONOMIQUE.
 PA Bolocrine A, Sorokine A, Renault P, Ehrlich SD;
 PI WPI; 2002-043418/06.
 DR New nucleotide sequence useful in the identification or Lactococcus
 XX lactis and related species -
 PT Claim 6; SEQ ID No 946; 2504bp; French.
 PS The present invention is related to a Lactococcus lactis nucleotide
 XX sequence (AB90521) and related proteins (AB53300-AB55621). The
 CC nucleic acid sequence is useful in the detection and/or amplification of
 CC nucleic acid sequence, particularly to identify Lactococcus lactis or
 CC related species. The proteins of the invention are useful for the
 CC biosynthesis or biodegradation of a composition of interest. The
 CC invention helps research in lactic bacteria, particularly useful in the
 CC production of yogurt and cheese.
 CC Note: The sequence data for this patent is based on equivalent patent
 CC WO200177334 (published 18-OCT-2001) which is available in electronic
 CC format directly from WIPO at ftp.wipo.int/pub/published_pat_sequences.
 XX
 SQ Sequence 306 AA;

Query Match 1.9%; Score 8; DB 23; Length 306;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 TLNAGDT 207
 |||||
 Db 202 TLNAGDT 209

RESULT 14
 ABP65884
 ID ABP65884 standard; Protein; 318 AA.
 AC ABP65884;
 XX
 DT 19-NOV-2002 (first entry)
 XX
 DE Bifidobacterium longum NCC2705 ORF amino acid sequence SEQ ID NO:628.
 XX
 KW Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;
 KW anti-diarrheic; antibacterial; inhibitor of Salmonella; detection;
 KW identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;
 XX rotavirus; food composition; pharmaceutical composition.
 OS Bifidobacterium longum.
 XX
 PN EP1227152-A1.
 XX
 PD 31-JUL-2002.
 XX
 PF 30-JAN-2001; 2001EP-0102050.
 XX
 PR 30-JAN-2001; 2001EP-0102050.
 XX
 PA (NEST) SOC PROD NESTLE SA.
 XX
 DR WPI; 2002-668397/72.
 XX
 PT Novel polynucleotide comprising Bifidobacterium genome sequence useful
 PT as a probe or primer for detecting and/or identifying Bifidobacterium
 PT longum in a biological sample -
 PS Claim 3; SEQ ID 628; 80bp; English.
 CC The present invention describes a polynucleotide (I) comprising a

CC sequence of a Bifidobacterium genome selected from the nucleotide
 CC sequences given in AB081842 and AB081843, or a sequence exhibiting at
 CC least 90% identity or which hybridises with the sequences given in
 CC AB081842 and AB081843. Also described is a polynucleotide (II) encoding
 CC a fusion protein, comprising a sequence selected from 1097 sequences
 CC given in ABP65258 to ABP66354 ligated in frame to a polynucleotide
 CC encoding a heterologous polypeptide. (I) has anti-diarrheic and
 CC antibacterial activities, and can be used as an inhibitor of Salmonella.
 CC (I) (which is a probe) is useful for the detection and/or identification
 CC of Bifidobacterium longum in a biological sample. A carrier containing
 CC the lactic acid bacterium Bifidobacterium longum NCC2705 (CNCM I-2618)
 CC can be used for preventing and/or treating diarrhoea brought about by
 CC pathogenic bacteria and/or rotavirus. The carrier is a food composition
 CC selected from milk, yogurt, curd, cheese, fermented milks, milk based
 CC fermented products, ice-creams, fermented cereal based products, milk
 CC based powders, infant formula, pet food or a pharmaceutical composition
 CC selected from tablets, liquid bacterial suspensions, dried oral
 CC supplement, wet oral supplement, dry tube feeding or wet tube feeding.
 CC (I) is useful in DNA arrays or chips to carry out analysis of the
 CC expression of the Bifidobacterium gene. AB081844 to AB081850 represent
 CC Bifidobacterium related nucleotide sequences given in the Sequence
 CC listing from the present invention but not mentioned further within the
 CC specification.
 CC N.B. The sequence data for this patent is not represented in the printed
 CC specification but is based on sequence information supplied by the
 CC European Patent Office.
 XX
 SQ Sequence 318 AA;

Query Match 1.9%; Score 8; DB 23; Length 318;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 AALAAAV 18
 |||||
 Db 210 AALAAAV 217

RESULT 15
 ABG28560
 ID ABG28560 standard; Protein; 426 AA.
 AC ABG28560;
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #28551.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 XX
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 DR Dmanac RT, Liu C, Tang YT;
 XX
 PT WPI; 2001-639362/73.
 DR N-PSDB; AAS92747.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -

CC This sequence corresponds to a human methionine aminopeptidase.
 CC Over-expression of the protein in a protein synthesis system (in vivo,
 CC in vitro or recombinant) is used to promote protein synthesis by removing
 CC N-terminal methionine from the desired protein. Methionine aminopeptidase
 CC and eIF-2 associated glycoprotein (p67) are shown to be substantially the
 CC same protein and so providing a cellular system with the cloned
 CC nucleotide sequence will serve both cellular functions.

XX
 SQ Sequence 478 AA;

Query Match 1.9%; Score 8; DB 20; Length 478;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 203 NAGDTTVL 210
 |||||
 Db 235 NAGDTTVL 242

RESULT 18
 ID AAM93215
 AC AAM93215 standard; Protein; 478 AA.

XX
 AC AAM93215;
 XX
 DT 27-MAY-1999 (first entry)

DE Human p67 homologue with methionine aminopeptidase activity protein.
 XX
 KW Methionine aminopeptidase; eIF-2; protein synthesis; N-terminal; p67;
 KW eukaryotic initiation factor-2 associated glycoprotein; regulatory;
 KW protein modification; N-myristoylation.

XX Homo sapiens.

XX US5885820-A.

XX 23-MAR-1999.

XX 18-MAR-1998; 98US-0040799.

XX 31-JAN-1996; 96US-0595025.

XX 18-MAR-1998; 98US-0040799.

XX (UYSL-) UNTV ST LOUIS.

XX Chang Y;

XX PI

XX DR WPI; 1999-228541/19.

XX DR N-PSDB; AAM22709, AAM22710.

PT Protein with methionine aminopeptidase activity - similar to
 PT eukaryotic initiation factor-2 associated glycoprotein is new

PS Claim 2; Fig 1; 11pp; English.

XX This sequence represents a human methionine aminopeptidase found to be
 CC similar to eukaryotic initiation factor-2 (eIF-2) associated
 CC glycoprotein (p67). The methionine aminopeptidase protein removes the
 CC N-terminal Met residue from proteins during protein synthesis. This is
 CC essential for subsequent modification of protein, such as in
 CC N-myristoylation. The methionine aminopeptidase is substantially similar
 CC to eukaryotic initiation factor-2 (eIF-2) associated glycoprotein (p67),
 CC and so may facilitate the function of an eukaryotic initiation factor,
 CC thus having a regulatory role in regulation of protein synthesis. The
 CC protein may facilitate protein synthesis by protecting eIF-2 from
 CC phosphorylation. The methionine aminopeptidase polynucleotides can
 CC be used to monitor synthesis of the protein peptidase.

XX Sequence 478 AA;

Query Match 1.9%; Score 8; DB 20; Length 478;
 Best Local Similarity 100.0%; Pred. No. 41;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 203 NAGDTTVL 210
 |||||
 Db 235 NAGDTTVL 242

RESULT 19
 ID AAM94763
 AC AAM94763 standard; protein; 478 AA.

XX
 AC AAM94763;
 XX
 DT 28-APR-1999 (first entry)

DE Mouse type 2 methionine aminopeptidase (MetAP2) putative sequence.

XX Ovalicin; fumagillin; type 2 methionine aminopeptidase; inhibitor;
 KW MetAP2; angiogenesis; tumour; diabetic retinopathy; arteriosclerosis;
 KW inflammatory disease; immune reaction; autoimmune disease; allergy;
 KW tissue graft rejection; mouse.

XX Mus sp.

XX OS

XX FH

XX Key

XX Location/Qualifiers

XX Misc-difference 418

XX /label= unknown

XX W09856372-A1.

XX 17-DEC-1998.

XX 08-JUN-1998; 96WO-US11775.

XX 09-JUN-1997; 97US-0049159.

XX (MASI) MASSACHUSETTS INST TECHNOLOGY.

XX Griffith EC, Liu JO, Su Z;

XX WPI; 1999-080848/07.

XX DR

XX PT New ovalicin and fumagillin derivatives - are inhibitors of type 2

XX PT methionine amino-peptidase, useful for treating or diagnosing

XX PT diseases involving abnormal angiogenesis or immune reactions

XX PS Disclosure; Fig 2; 99pp; English.

XX The invention relates to ovalicin and fumagillin derivatives that can

XX CC inhibit type 2 methionine aminopeptidase (MetAP2). These inhibitors are

XX CC useful for treating and/or diagnosing diseases involving abnormal

XX CC angiogenesis (e.g. tumours, diabetic retinopathy, inflammatory diseases,

XX CC arteriosclerosis) or immune reactions which result in pathology (e.g.

XX CC autoimmune disease, allergy and tissue graft rejection). The present

XX CC sequence represents a putative amino acid sequence of mouse MetAP2.

XX Sequence 478 AA;

Query Match 1.9%; Score 8; DB 20; Length 478;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 203 NAGDTTVL 210
 |||||
 Db 235 NAGDTTVL 242

RESULT 20
 ID AAM94764
 AC AAM94764 standard; protein; 478 AA.

XX AAM94764;

XX

DT 28-APR-1999 (first entry)
 XX Rat type 2 methionine aminopeptidase (MetAP2).
 DE
 XX
 XX Ovaliclin; fumagillin; type 2 methionine aminopeptidase; inhibitor;
 KW MetAP2; angiogenesis; tumour; diabetic retinopathy; arteriosclerosis;
 KW inflammatory disease; immune reaction; autoimmune disease; allergy;
 KW tissue graft rejection; rat.
 XX
 OS Rattus sp.
 XX
 XX WO9856372-A1.
 XX
 XX 17-DEC-1998.
 XX
 XX 08-JUN-1998; 98WO-US11775.
 XX
 XX 09-JUN-1997; 97US-0049159.
 XX
 XX (MASI) MASSACHUSETTS INST TECHNOLOGY.
 XX
 XX Griffith EC, Liu JO, Su Z;
 XX
 XX WPI; 1999-080848/07.
 XX
 XX
 XX New ovaliclin and fumagillin derivatives - are inhibitors of type 2
 PT methionine amino-peptidase, useful for treating or diagnosing
 PT diseases involving abnormal angiogenesis or immune reactions
 XX
 XX Disclosure; Fig 2; 99pp; English.
 XX
 XX The invention relates to ovaliclin and fumagillin derivatives that can
 CC inhibit type 2 methionine aminopeptidase (MetAP2). These inhibitors are
 CC useful for treating and/or diagnosing diseases involving abnormal
 CC angiogenesis (e.g. tumours, diabetic retinopathy, inflammatory diseases,
 CC arteriosclerosis) or immune reactions which result in pathology (e.g.
 CC autoimmune disease, allergy and tissue graft rejection). The present
 CC sequence represents the amino acid sequence of rat MetAP2.
 CC
 SQ Sequence 478 AA;
 Query Match 1.9%; Score 8; DB 20; Length 478;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 203 NAGDTTVL 210
 |||||
 Db 235 NAGDTTVL 242
 RESULT 21
 AAM94765
 ID AAM94765 standard; protein; 478 AA.
 XX
 XX AAM94765;
 XX
 XX 28-APR-1999 (first entry)
 XX
 XX Human type 2 methionine aminopeptidase (MetAP2).
 DE
 XX Ovaliclin; fumagillin; type 2 methionine aminopeptidase; inhibitor;
 KW MetAP2; angiogenesis; tumour; diabetic retinopathy; arteriosclerosis;
 KW inflammatory disease; immune reaction; autoimmune disease; allergy;
 KW tissue graft rejection; human.
 XX
 XX Homo sapiens.
 OS
 XX
 XX WO9856372-A1.
 XX
 XX 17-DEC-1998.
 XX
 XX 08-JUN-1998; 98WO-US11775.
 XX
 XX

PR 09-JUN-1997; 97US-0049159.
 XX
 XX (MASI) MASSACHUSETTS INST TECHNOLOGY.
 XX
 XX Griffith EC, Liu JO, Su Z;
 XX
 XX WPI; 1999-080848/07.
 XX
 XX
 XX New ovaliclin and fumagillin derivatives - are inhibitors of type 2
 PT methionine amino-peptidase, useful for treating or diagnosing
 PT diseases involving abnormal angiogenesis or immune reactions
 XX
 XX Disclosure; Fig 2; 99pp; English.
 XX
 XX The invention relates to ovaliclin and fumagillin derivatives that can
 CC inhibit type 2 methionine aminopeptidase (MetAP2). These inhibitors are
 CC useful for treating and/or diagnosing diseases involving abnormal
 CC angiogenesis (e.g. tumours, diabetic retinopathy, inflammatory diseases,
 CC arteriosclerosis) or immune reactions which result in pathology (e.g.
 CC autoimmune disease, allergy and tissue graft rejection). The present
 CC sequence represents the amino acid sequence of human MetAP2.
 CC
 SQ Sequence 478 AA;
 Query Match 1.9%; Score 8; DB 20; Length 478;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 203 NAGDTTVL 210
 |||||
 Db 235 NAGDTTVL 242
 RESULT 22
 ABB50275
 ID ABB50275 standard; Protein; 478 AA.
 XX
 XX ABB50275;
 XX
 XX 08-FEB-2002 (first entry)
 XX
 XX eIF-2-associated p67 ovarian tumour marker protein, SEQ ID NO:39.
 DE
 XX Ovarian tumour marker gene; human; overexpression; upregulation;
 KW epithelial tumour; cancer; diagnosis; prognosis; disease monitoring;
 KW identification; serous cystadenoma; borderline serous tumour;
 KW serous cystadenocarcinoma; mucinous cystadenocarcinoma;
 KW mucinous cystadenoma; borderline mucinous tumour; endometrioid carcinoma;
 KW undifferentiated carcinoma; clear cell adenocarcinoma; cystadenofibroma;
 KW adenofibroma; Brenner tumour; serial analysis of gene expression; SAGE;
 KW immune response pathway; cell proliferation regulation; protein folding;
 KW membrane localised; secreted; therapeutic target; cytostatic;
 KW gene therapy; vaccine.
 XX
 XX Homo sapiens.
 OS
 XX
 XX WO200175177-A2.
 XX
 XX 11-OCT-2001.
 XX
 XX 03-APR-2001; 2001WO-US10947.
 XX
 XX 03-APR-2000; 2000US-194336P.
 XX
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX
 XX Morin PJ, Sherman-Baust CA, Pizer ES, Hough CD;
 XX
 XX WPI; 2001-626450/72.
 XX
 XX N-P8DB; ABA83100.
 XX
 XX
 XX Detecting and identifying ovarian tumor, identifying increased risk for
 PT developing ovarian cancer, and determining effectiveness of ovarian

PT cancer treatment, by measuring expression level of ovarian tumor marker gene -

PS Claim 22; Page 87-88; 140pp; English.

XX The invention relates to methods for diagnosing and prognosing ovarian
CC tumors in an individual via the detection and measurement of the
CC expression of ovarian tumor marker genes (ABA83081-ABA83122, ABA83180,
CC ABA83182 and ABA83184) or segments thereof (ABA83123-ABA83169, ABA83179,
CC ABA83181 and ABA83183). The methods of the invention are useful for
CC detecting an ovarian tumor in a patient, for identifying an individual
CC at increased risk for developing ovarian cancer, in prognostic tests for
CC assessing the relative severity of ovarian cancer, in tests for
CC monitoring a patient in remission from ovarian cancer and in tests for
CC monitoring disease status in a patient being treated for ovarian cancer.
CC The methods can additionally be used to identify a particular tumor as
CC being an ovarian tumor (i.e., an epithelial ovarian tumor selected from
CC serous cystadenoma, borderline serous tumor, serous cystadenocarcinoma,
CC mucinous cystadenoma, borderline mucinous tumor, mucinous
CC cystadenocarcinoma, endometrioid carcinoma, undifferentiated carcinoma,
CC clear cell adenocarcinoma, cystadenofibroma, adenofibroma and Brenner
CC tumor. The ovarian tumor marker genes of the invention were identified
CC using SAGE (serial analysis of gene expression) and were found to be
CC overexpressed in a broad variety of ovarian epithelial tumor cells
CC relative to normal ovarian epithelial cells. The marker genes are
CC implicated in immune response pathways, in the regulation of cell
CC proliferation and in protein folding, and many of these are membrane-
CC localised or secreted. In addition to their use as diagnostic and
CC prognostic markers, the ovarian tumor marker genes or their encoded
CC proteins may be used as therapeutic targets for the treatment and
CC prevention of ovarian cancer. Sequences ABB50257-ABB50299 represent
CC proteins encoded by ovarian tumor marker genes of the invention.

XX Sequence 478 AA;

Query Match 1.9%; Score 8; DB 22; Length 478;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 203 NAGDTTVL 210
DB 235 NAGDTTVL 242

RESULT 23
AAB28377
ID AAB28377 standard; Protein; 478 AA.

XX AAB28377;

AC 16-FEB-2001 (first entry)

DT Human methionine aminopeptidase 2.

XX Antiinflammatory; cytosolic; antibacterial; methionine aminopeptidase 2;

KW inhibitor; MetAP2; eukaryotic initiation factor associated protein; p67;

KW eIF-2; protein synthesis; antisense oligonucleotide; infection; human;

XX inflammation; tumor.

OS Homo sapiens.

XX US6136604-A.

XX 24-OCT-2000.

XX 27-OCT-1999; 99US-0428584.

XX 27-OCT-1999; 99US-0428584.

XX (ISIS-) ISIS PHARM INC.

XX Monia BP, Wyatt J;

DR WPI; 2001-030942/04.

XX N-PSDB; AAC67683.

PT New antisense compounds which specifically hybridize with and inhibit
PT human methionine aminopeptidase 2 expression, useful for treating
PT methionine aminopeptidase 2 related disorders and preventing
PT inflammation or tumor formation -

PS Disclosure; Columns 45-50; 39pp; English.

XX The present sequence is human methionine aminopeptidase 2 (also known as
CC MetAP2 and eukaryotic initiation factor (eIF-2) associated protein,
CC p67). MetAP2 is a cellular glycoprotein that promotes protein synthesis
CC in the presence of active eIF-2 kinases by protecting the eIF-2 alpha
CC subunit from phosphorylation. The present invention relates to antisense
CC oligonucleotides (AAC67690-C67767) which inhibit expression of the
CC present sequence. The antisense oligonucleotides of the present invention
CC may be used for treating a patient suspected of having or being prone to
CC a disease or condition associated with expression of MetAP2. The
CC antisense oligonucleotides may further be used prophylactically, e.g. to
CC prevent or delay infection, inflammation or tumor formation.

XX Sequence 478 AA;

Query Match 1.9%; Score 8; DB 22; Length 478;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 203 NAGDTTVL 210
DB 235 NAGDTTVL 242

RESULT 24
ABG76374
ID ABG76374 standard; protein; 478 AA.

XX ABG76374;

AC 20-MAY-2003 (first entry)

DT Human type 2 aminopeptidase (MetAP2) variant dnMetAP2.

XX Human; dominant negative variant; type 2 methionine aminopeptidase;

KW MetAP2; translation domain; fungal infection; cell proliferation;

KW p53 function; immune system; angiogenesis; opportunistic infection;

XX cancer; cytostatic; fungicide; immunomodulatory; anti-angiogenic;

XX variant; dnMetAP2.

OS Homo sapiens.

XX Synthetic.

XX Key

XX Location/Qualifiers

XX Misc-difference 219

XX /note= "Any naturally occurring amino acid"

XX Misc-difference 221

XX /note= "Any amino acid, except His"

XX Misc-difference 251

XX /note= "Any naturally occurring amino acid"

XX Misc-difference 262

XX /note= "Any naturally occurring amino acid"

XX Misc-difference 328

XX /note= "Any naturally occurring amino acid"

XX Misc-difference 331

XX /note= "Any naturally occurring amino acid"

XX Misc-difference 338..339

XX /note= "Any naturally occurring amino acid"

XX Misc-difference 364

XX /note= "Any naturally occurring amino acid"

XX Misc-difference 444

XX /note= "Any naturally occurring amino acid"

XX Misc-difference 447

XX /note= "Any naturally occurring amino acid"

XX Misc-difference 447

XX /note= "Any naturally occurring amino acid"

```

FT Misc-difference 459 /note= "Any naturally occurring amino acid"
XX
XX US2002182701-A1.
XX
XX 05-DEC-2002.
XX
XX 30-AUG-2001; 2001US-0943123.
XX
XX 30-AUG-2001; 2001US-0943123.
XX
XX (UYSL-) UNIV SAINT LOUIS.
XX
XX Chang Y, Micka WS, Vetro JA;
XX
XX WPI; 2003-328620/31.
XX
XX
XX New variant type 2 methionine aminopeptidase polypeptide, useful for
XX preparing a medicament for treating a disease mediated by fungal
XX infection, cell proliferation, decreased function of p53, immune system
XX activity or angiogenesis -
XX
XX Claim 3; Page 16-17; 46pp; English.
XX
XX The present invention relates to dominant negative variants of
XX type 2 methionine aminopeptidase (MetAP2) containing a translation
XX domain. The MetAP2 variant polypeptides are useful for preparing a
XX medicament for treating a disease mediated by fungal infection, cell
XX proliferation, decreased function of p53, immune system activity,
XX or preferably angiogenesis. The MetAP2 polypeptides of the
XX invention may be used for treating subjects suffering from cancer,
XX diseases mediated by the immune system or opportunistic infections
XX using inhibitors of MetAP2. The present sequence represents
XX human MetAP2 variant dnMetAP2.
XX
XX Sequence 478 AA;
SQ
Query Match 1.9%; Score 8; DB 24; Length 478;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 203 NAGDTTVL 210
Db 235 NAGDTTVL 242
RESULT 25
ABG76375 standard; protein; 478 AA.
XX
XX ABG76375;
XX
XX 20-MAY-2003 (first entry)
XX
XX Mouse type 2 aminopeptidase (MetAP2) variant.
XX
XX Mouse; dominant negative variant; type 2 methionine aminopeptidase;
XX MetAP2; translation domain; fungal infection; cell proliferation;
XX p53 function; immune system; angiogenesis; opportunistic infection;
XX cancer; cytostatic; fungicide; immunomodulatory; anti-angiogenic;
XX variant.
XX
XX Mus sp.
XX Synthetic.
XX
XX Key Location/Qualifiers
XX Misc-difference 219 /note= "Any naturally occurring amino acid"
XX
XX Misc-difference 231 /note= "Any amino acid, except His"
XX
XX Misc-difference 251 /note= "Any naturally occurring amino acid"
XX
XX Misc-difference 262

```

```

FT /note= "Any naturally occurring amino acid"
FT Misc-difference 328 /note= "Any naturally occurring amino acid"
FT Misc-difference 331 /note= "Any naturally occurring amino acid"
FT Misc-difference 331 /note= "Any naturally occurring amino acid"
FT Misc-difference 338..339 /note= "Any naturally occurring amino acid"
FT Misc-difference 364 /note= "Any naturally occurring amino acid"
FT Misc-difference 364 /note= "Any naturally occurring amino acid"
FT Misc-difference 444 /note= "Any naturally occurring amino acid"
FT Misc-difference 447 /note= "Any naturally occurring amino acid"
FT Misc-difference 447 /note= "Any naturally occurring amino acid"
FT Misc-difference 459 /note= "Any naturally occurring amino acid"
FT Misc-difference 459 /note= "Any naturally occurring amino acid"
XX
XX US2002182701-A1.
XX
XX 05-DEC-2002.
XX
XX 30-AUG-2001; 2001US-0943123.
XX
XX 30-AUG-2001; 2001US-0943123.
XX
XX (UYSL-) UNIV SAINT LOUIS.
XX
XX Chang Y, Micka WS, Vetro JA;
XX
XX WPI; 2003-328620/31.
XX
XX New variant type 2 methionine aminopeptidase polypeptide, useful for
XX preparing a medicament for treating a disease mediated by fungal
XX infection, cell proliferation, decreased function of p53, immune system
XX activity or angiogenesis -
XX
XX Claim 3; Page 17-18; 46pp; English.
XX
XX The present invention relates to dominant negative variants of
XX type 2 methionine aminopeptidase (MetAP2) containing a translation
XX domain. The MetAP2 variant polypeptides are useful for preparing a
XX medicament for treating a disease mediated by fungal infection, cell
XX proliferation, decreased function of p53, immune system activity,
XX or preferably angiogenesis. The MetAP2 polypeptides of the
XX invention may be used for treating subjects suffering from cancer,
XX diseases mediated by the immune system or opportunistic infections
XX using inhibitors of MetAP2. The present sequence represents a
XX mouse MetAP2 variant.
XX
XX Sequence 478 AA;
SQ
Query Match 1.9%; Score 8; DB 24; Length 478;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 203 NAGDTTVL 210
Db 235 NAGDTTVL 242
RESULT 26
ABG76377 standard; protein; 478 AA.
XX
XX ABG76377;
XX
XX 20-MAY-2003 (first entry)
XX
XX Human type 2 aminopeptidase (MetAP2).
XX
XX Human; dominant negative variant; type 2 methionine aminopeptidase;
XX MetAP2; translation domain; fungal infection; cell proliferation;
XX p53 function; immune system; angiogenesis; opportunistic infection;
XX cancer; cytostatic; fungicide; immunomodulatory; anti-angiogenic;
XX

```

KW enzyme.
 XX Homo sapiens.
 OS
 XX US2002182701-A1.
 PN
 XX 05-DEC-2002.
 PD
 XX 30-AUG-2001; 2001US-0943123.
 PF
 XX 30-AUG-2001; 2001US-0943123.
 PR
 XX (UYSL-) UNIV SAINT LOUIS.
 PA
 XX Chang Y, Micka WS, Vetro JA;
 PI
 XX WPI; 2003-328620/31.
 DR
 XX New variant type 2 methionine aminopeptidase polypeptide, useful for
 PT preparing a medicament for treating a disease mediated by fungal
 PT infection, cell proliferation, decreased function of p53, immune system
 PT activity or angiogenesis -
 CC
 XX Claim 5; Fig 1; 46pp; English.
 PS
 XX The present invention relates to dominant negative variants of
 CC type 2 methionine aminopeptidase (MetAP2) containing a translation
 CC domain. The MetAP2 variant polypeptides are useful for preparing a
 CC medicament for treating a disease mediated by fungal infection, cell
 CC proliferation, decreased function of p53, immune system activity,
 CC or preferably angiogenesis. The MetAP2 polypeptides of the
 CC invention may be used for treating subjects suffering from cancer,
 CC diseases mediated by the immune system or opportunistic infections
 CC using inhibitors of MetAP2. The present sequence represents human
 CC MetAP2.
 CC
 SQ Sequence 478 AA;
 XX
 XX Query Match 1.9%; Score 8; DB 24; Length 478;
 XX Best Local Similarity 100.0%; Pred. No. 41;
 XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 203 NAGDTTVL 210
 DB 235 NAGDTTVL 242
 XX
 XX RESULT 27
 XX ABG76378
 XX ID ABG76378 standard; protein; 478 AA.
 XX AC
 XX ABG76378;
 XX
 XX 20-MAY-2003 (first entry)
 DT
 XX
 XX Mouse type 2 aminopeptidase (MetAP2).
 DE
 XX
 XX Mouse; dominant negative variant; type 2 methionine aminopeptidase;
 KW MetAP2; translation domain; fungal infection; cell proliferation;
 KW p53 function; immune system; angiogenesis; opportunistic infection;
 KW cancer; cyostatic; fungicide; immunomodulatory; anti-angiogenic;
 KW enzyme.
 XX
 XX Mus sp.
 OS
 XX US2002182701-A1.
 PN
 XX 05-DEC-2002.
 PD
 XX 30-AUG-2001; 2001US-0943123.
 PF
 XX 30-AUG-2001; 2001US-0943123.
 PR
 XX (UYSL-) UNIV SAINT LOUIS.
 PA
 XX Chang Y, Micka WS, Vetro JA;
 PI
 XX WPI; 2003-328620/31.
 DR
 XX New variant type 2 methionine aminopeptidase polypeptide, useful for
 PT preparing a medicament for treating a disease mediated by fungal
 PT infection, cell proliferation, decreased function of p53, immune system
 PT activity or angiogenesis -
 CC
 XX Claim 5; Fig 1; 46pp; English.
 PS
 XX The present invention relates to dominant negative variants of
 CC type 2 methionine aminopeptidase (MetAP2) containing a translation
 CC domain. The MetAP2 variant polypeptides are useful for preparing a
 CC medicament for treating a disease mediated by fungal infection, cell
 CC proliferation, decreased function of p53, immune system activity,
 CC or preferably angiogenesis. The MetAP2 polypeptides of the
 CC invention may be used for treating subjects suffering from cancer,
 CC diseases mediated by the immune system or opportunistic infections
 CC using inhibitors of MetAP2. The present sequence represents human
 CC MetAP2.
 CC
 SQ Sequence 478 AA;
 XX
 XX Query Match 1.9%; Score 8; DB 24; Length 478;
 XX Best Local Similarity 100.0%; Pred. No. 41;
 XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

PA (UYSL-) UNIV SAINT LOUIS.
 XX
 XX Chang Y, Micka WS, Vetro JA;
 PI
 XX WPI; 2003-328620/31.
 DR
 XX New variant type 2 methionine aminopeptidase polypeptide, useful for
 PT preparing a medicament for treating a disease mediated by fungal
 PT infection, cell proliferation, decreased function of p53, immune system
 PT activity or angiogenesis -
 CC
 XX Disclosure; Fig 1; 46pp; English.
 PS
 XX The present invention relates to dominant negative variants of
 CC type 2 methionine aminopeptidase (MetAP2) containing a translation
 CC domain. The MetAP2 variant polypeptides are useful for preparing a
 CC medicament for treating a disease mediated by fungal infection, cell
 CC proliferation, decreased function of p53, immune system activity,
 CC or preferably angiogenesis. The MetAP2 polypeptides of the
 CC invention may be used for treating subjects suffering from cancer,
 CC diseases mediated by the immune system or opportunistic infections
 CC using inhibitors of MetAP2. The present sequence represents mouse
 CC MetAP2.
 CC
 SQ Sequence 478 AA;
 XX
 XX Query Match 1.9%; Score 8; DB 24; Length 478;
 XX Best Local Similarity 100.0%; Pred. No. 41;
 XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 203 NAGDTTVL 210
 DB 235 NAGDTTVL 242
 XX
 XX RESULT 28
 XX ABU07490
 XX ID ABU07490 standard; Protein; 478 AA.
 XX AC
 XX ABU07490;
 XX
 XX 28-JAN-2003 (first entry)
 DT
 XX
 XX Protein differentially regulated in prostate cancer #93.
 DE
 XX
 XX Prostate cancer; gene expression; differential regulation;
 KW molecular marker; drug target; cancer detection; cancer diagnosis;
 KW cancer staging; cancer grading; cancer assessing; cancer monitoring.
 KW
 XX Homo sapiens.
 OS
 XX WO200281638-A2.
 PN
 XX 17-OCT-2002.
 PD
 XX
 XX 08-APR-2002; 2002WO-US10824.
 PF
 XX
 XX 06-APR-2001; 2001US-281731P.
 PR
 XX 06-APR-2001; 2001US-281732P.
 PA
 XX (ORIG-) ORIGENE TECHNOLOGIES INC.
 XX
 XX Sun Z, Jay G;
 PI
 XX WPI; 2003-058520/05.
 DR
 XX N-PSDB; ABX10395.
 DR
 XX Novel genes which are differentially regulated in prostate cancer,
 PT useful for diagnosing prostate cancer in prostate tissue sample and
 PT assessing therapeutic or preventive intervention in prostate cancer
 PT patients -
 CC
 XX Claim 1; Page 398-399; 416pp; English.
 PS

XX The invention describes genes (I) which are differentially regulated in prostate cancer. (I) is useful for diagnosing a prostate cancer in a sample comprising prostate tissue, which involves determining the number of target genes which are differentially-regulated in the sample, where the number is indicative of the probability that the sample comprises prostate cancer. (I) is useful for assessing a therapeutic or preventive intervention in a subject having a prostate cancer, which involves determining the expression levels in a sample comprising prostate tissue of target genes which are differentially-regulated in prostate cancer. Preferably, the expression levels of at least 10 genes are determined. (I) is also useful for identifying agents that modulate a biological activity of a polypeptide differentially-regulated in prostate cancer cells, which involves contacting a polypeptide differentially-regulated in prostate cancer cells with a test agent under conditions effective for the test agent to modulate a biological activity of the polypeptide, and determining whether the test agent modulates the biological activity. (I) is useful as molecular markers, as drug targets, and for detecting, diagnosing, staging, grading, assessing, monitoring, prognosticating, preventing or treating, determining predisposition to diseases and conditions especially relating to prostate cancer. (I) and its expression products are used in the diagnostic test to assay for presence of cancer e.g., in tissue sections, in biopsy sample, in total RNA, in lymph, in blood etc. (I) is useful for assessing cancer e.g., to determine the type of cancer, its stage of development, the nature of genetic defect, etc. The polypeptide encoded by (I) can be used as target for therapy or drug discovery. (I) can also be used for expressing the polypeptide and thus for searching specific binding partners of the polypeptide. (I) is useful in therapeutic applications to treat prostate cancer. The identification of specific genes, and groups of genes, expressed in pathways physiologically relevant to prostate cancer permits the definition of functional and disease pathways and the delineation of targets in these pathways which are useful in diagnostic, therapeutic, and clinical applications. This is the amino acid sequence of a protein differentially regulated in prostate cancer.

XX Sequence 478 AA;

Query Match 1.9%; Score 8; DB 24; Length 478;

Best Local Similarity 100.0%; Pred. No. 41; Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 203 NAGDTTVL 210
|||||||
DB 235 NAGDTTVL 242

RESULT 29

ABG76381

ID ABG76381 standard; protein; 480 AA.

XX AC ABG76381;

XX DT 20-MAY-2003 (first entry)

XX DE Rat type 2 aminopeptidase (MeAP2) variant dnmMeAP2.

XX Rat; dominant negative variant; type 2 methionine aminopeptidase; MeAP2; translation domain; fungal infection; cell proliferation; p53 function; immune system; angiogenesis; opportunistic infection; cancer; cytosolic; fungicide; immunomodulatory; anti-angiogenic; variant; dnmMeAP2.

XX OS Rattus sp.

XX OS Synthetic.

XX Key Location/Qualifiers

XX FT Misc-difference 219

XX FT /note= "Any naturally occurring amino acid"

XX FT Misc-difference 231

XX FT /note= "Any amino acid, except His"

XX FT Misc-difference 251

XX FT /note= "Any naturally occurring amino acid"

FT Misc-difference 262
FT /note= "Any naturally occurring amino acid"
FT Misc-difference 328
FT /note= "Any naturally occurring amino acid"
FT Misc-difference 331
FT /note= "Any naturally occurring amino acid"
FT Misc-difference 338..339
FT /note= "Any naturally occurring amino acid"
FT Misc-difference 364
FT /note= "Any naturally occurring amino acid"
FT Misc-difference 444
FT /note= "Any naturally occurring amino acid"
FT Misc-difference 447
FT /note= "Any naturally occurring amino acid"
FT Misc-difference 459
FT /note= "Any naturally occurring amino acid"

PN US2002182701-A1.

XX 05-DEC-2002.

XX 30-AUG-2001; 2001US-0943123.

XX 30-AUG-2001; 2001US-0943123.

XX (USL-) UNIV SAINT LOUIS.

XX Chang Y, Micka WS, Vetro JA;

XX WPI; 2003-328620/31.

PT New variant type 2 methionine aminopeptidase polypeptide, useful for preparing a medicament for treating a disease mediated by fungal infection, cell proliferation, decreased function of p53, immune system activity or angiogenesis -

XX Claim 3; Page 28-29; 46pp; English.

XX The present invention relates to dominant negative variants of type 2 methionine aminopeptidase (MeAP2) containing a translation domain. The MeAP2 variant polypeptides are useful for preparing a medicament for treating a disease mediated by fungal infection, cell proliferation, decreased function of p53, immune system activity, or preferably angiogenesis. The MeAP2 polypeptides of the invention may be used for treating subjects suffering from cancer, diseases mediated by the immune system or opportunistic infections CC using inhibitors of MeAP2. The present sequence represents rat MeAP2 variant dnmMeAP2.

XX Sequence 480 AA;

Query Match 1.9%; Score 8; DB 24; Length 480;

Best Local Similarity 100.0%; Pred. No. 41; Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 203 NAGDTTVL 210
|||||||
DB 235 NAGDTTVL 242

RESULT 30

ABG76382

ID ABG76382 standard; protein; 480 AA.

XX AC ABG76382;

XX DT 20-MAY-2003 (first entry)

XX DE Rat type 2 aminopeptidase (MeAP2).

XX Rat; dominant negative variant; type 2 methionine aminopeptidase; MeAP2; translation domain; fungal infection; cell proliferation; p53 function; immune system; angiogenesis; opportunistic infection;

KW cancer; cytostatic; fungicide; immunomodulatory; anti-angiogenic;
 KW enzyme.
 XX Rattus sp.
 OS US2002182701-A1.
 XX
 XX
 PD 05-DEC-2002.
 XX
 PF 30-AUG-2001; 2001US-0943123.
 XX
 PR 30-AUG-2001; 2001US-0943123.
 XX
 PA (US&L-) UNIV SAINT LOUIS.
 XX
 PI Chang Y, Micka WS, Vetro JA;
 DR WPI; 2003-328620/31.
 XX
 PT New variant type 2 methionine aminopeptidase polypeptide, useful for
 PT preparing a medicament for treating a disease mediated by fungal
 PT infection, cell proliferation, decreased function of p53, immune system
 PT activity or angiogenesis -
 XX
 PS Disclosure; Fig 1; 46pp; English.
 XX
 CC The present invention relates to dominant negative variants of
 CC type 2 methionine aminopeptidase (MetAP2) containing a translation
 CC domain. The MetAP2 variant polypeptides are useful for preparing a
 CC medicament for treating a disease mediated by fungal infection, cell
 CC proliferation, decreased function of p53, immune system activity,
 CC or preferably angiogenesis. The MetAP2 polypeptides of the
 CC invention may be used for treating subjects suffering from cancer,
 CC diseases mediated by the immune system or opportunistic infections
 CC using inhibitors of MetAP2. The present sequence represents rat
 CC MetAP2.
 CC
 SO Sequence 480 AA;
 Query Match 1.9%; Score 8; DB 24; Length 480;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 203 NAGDTTVL 210
 DB 235 NAGDTTVL 242
 RESULT 31
 ID AAG73991 standard; Protein; 500 AA.
 XX
 AC AAG73991;
 XX
 DT 03-SEP-2001 (first entry)
 XX
 DE Human colon cancer antigen protein SEQ ID NO:4755.
 XX
 KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
 KW colorectal carcinoma.
 XX
 OS Homo sapiens.
 XX
 PN WO200122920-A2.
 XX
 PD 05-APR-2001.
 XX
 PF 28-SEP-2000; 2000WO-US26524.
 XX
 PR 29-SEP-1999; 99US-0157137.
 PR 03-NOV-1999; 99US-0163280.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.

XX
 PI Ruben SM, Barash SC, Birse CE, Rosen CA;
 XX
 DR WPI; 2001-235357/24.
 DR N-PSDB; AAH33422.
 XX
 PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
 PT useful for preventing, diagnosing and/or treating colorectal cancers -
 XX
 PS Claim 11; Page 6553-6555; 9803pp; English.
 XX
 CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have cytostatic activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing P.
 CC inactive proteins or to supplement the patient's own production of P.
 CC Additionally, N may be used to produce the colon cancer-associated Ps,
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
 CC and AAB77789 represent sequences used in the exemplification of the
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.
 XX
 SO Sequence 500 AA;
 Query Match 1.9%; Score 8; DB 22; Length 500;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 203 NAGDTTVL 210
 DB 257 NAGDTTVL 264
 RESULT 32
 ID ABG17833 standard; Protein; 545 AA.
 XX
 AC ABG17833;
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #17824.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS82020.
 XX

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
PS Claim 20; SEQ ID No 48192; 103pp; English.
XX
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 545 AA;
Query Match 1.9%; Score 8; DB 22; Length 545;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 10 RAALAAA 17
Db 234 RAALAAA 241
RESULT 33
ABG13539
ID ABG13539 standard; Protein; 609 AA.
XX
XX
AC ABG13539;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #13530.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX
PA (HYSE-) HYSEQ INC.
PI Drmanac RT, Liu C, Tang YT;
DR WPI; 2001-639362/73.
DR N-PSDB; AAS77726.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
PS Claim 20; SEQ ID No 43898; 103pp; English.
XX
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 609 AA;
Query Match 1.9%; Score 8; DB 22; Length 609;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 385 WASSENO 392
Db 537 WASSENO 544
RESULT 34
ABG30067
ID ABG30067 standard; Protein; 689 AA.
XX
XX
AC ABG30067;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #30068.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX
PA (HYSE-) HYSEQ INC.
PI Drmanac RT, Liu C, Tang YT;
DR WPI; 2001-639362/73.
DR N-PSDB; AAS94254.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

XX Claim 20; SEQ ID No 60426; 103pp; English.
PS
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 689 AA;
Query Match 1.9%; Score 8; DB 22; Length 689;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 47 VGVVTVHP 54
| | | | |
Db 290 VGVVTVHP 297
RESULT 35
AAU34554
ID AAU34554 standard; Protein; 891 AA.
XX
AC AAU34554;
XX
DT 14-FEB-2002 (first entry)
XX
DE E. coli cellular proliferation protein #135.
XX
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
OS Escherichia coli.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR N-PSDB; AAS52413.
XX

PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX
PS Example 3; Seq ID No 10147; 511pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 891 AA;
Query Match 1.9%; Score 8; DB 22; Length 891;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 10 RAALAAA 17
| | | | |
Db 36 RAALAAA 43
RESULT 36
AAU38250
ID AAU38250 standard; Protein; 892 AA.
XX
AC AAU38250;
XX
DT 14-FEB-2002 (first entry)
XX
DE Salmonella typhi cellular proliferation protein #141.
XX
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
OS Salmonella typhi.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR N-PSDB; AAS56109.
XX

PT New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 PS Example 3; Seq ID No 13843; 511pp; English.

CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

CC Sequence 892 AA;

Query Match 1.9%; Score 8; DB 22; Length 892;

Best Local Similarity 100.0%; Pred. No. 76;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 RAALALAA 17
 |||||
 DB 36 RAALALAA 43

RESULT 37

ABB61691
 ID ABB61691 standard; Protein; 968 AA.

AC ABB61691;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 11865.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

XX Drosophila melanogaster.

OS WO200171042-A2.

PN 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

DR WPI; 2001-656860/75.

XX N-PSDB; ABL05794.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PS Disclosure; SEQ ID NO 11865; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent.
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins
 CC (AB57737-AB572072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

CC Sequence 968 AA;

Query Match 1.9%; Score 8; DB 22; Length 968;

Best Local Similarity 100.0%; Pred. No. 82;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 117 AQLATQA 124
 |||||
 DB 111 AQLATQA 118

RESULT 38

ABB68075
 ID ABB68075 standard; Protein; 1300 AA.

AC ABB68075;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 31017.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

XX Drosophila melanogaster.

OS WO200171042-A2.

PN 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

DR WPI; 2001-656860/75.

XX N-PSDB; ABL12178.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PS Disclosure; SEQ ID NO 31017; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins
 CC (AB57737-AB572072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

CC Sequence 1300 AA;

Query Match 1.9%; Score 8; DB 22; Length 1300;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 RAAALAAA 17
 |||||
 DB 877 RAAALAAA 884

RESULT 39

ABG05084
 ID ABG05084 standard; Protein; 1882 AA.

XX ABG05084;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #5075.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.

DR N-PSDB; AAS69271.

XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -

PS Claim 20; SEQ ID NO 35443; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations in
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1882 AA;

Query Match 1.9%; Score 8; DB 22; Length 1882;

Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 QIABGKLL 243
 |||||
 DB 1729 QIABGKLL 1736

RESULT 40

AAM40183
 ID AAM40183 standard; Protein; 1883 AA.

XX AAM40183;

DT 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 3328.

XX Human; nocotropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.

OS Homo sapiens.

PN WO200153312-A1.

PD 26-JUL-2001.

PF 26-DEC-2000; 2000WO-US34263.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0596042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou F, Goodrich R, Drmanac RT;

DR WPI: 2001-442253/47.

DR N-PSDB; AAI59339.

XX Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -

PT Example 5; SEQ ID NO 3328; 10078bp; English.

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nocotropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S. disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

XX

SQ Sequence 1883 AA;

Query Match 1.9%; Score 8; DB 22; Length 1883;

Best Local Similarity 100.0%; Pred. No. 1.6e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 QIAEGKLL 243

|||||

Db 1730 QIAEGKLL 1737

Search completed: September 8, 2003, 14:08:37
 UOB Time : 62 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2003, 13:56:26 ; Search time 41 Seconds
(without alignments)
2593.114 Million cell updates/sec

Title: US-09-889-756A-2

Perfect score: 2019 1 MAFYAFKRAMRAALAAVAL.....AAPQSGVQTASEAKTASEAE 412

Sequence:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriaph:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	DB ID	Description
1	1993	98.7	412	16	Q9J50 neisseria m
2	1989	98.5	412	16	Q9J66 neisseria m
3	1278	63.3	271	2	Q51007 neisseria g
4	836	41.4	398	16	Q8Y39 raietonia s
5	777	38.5	385	16	Q8ZLN5 salmonella
6	770	38.1	388	16	Q9AA03 caulobacter
7	765	37.9	399	2	Q8P7C8 xanthomonas
8	763.5	37.8	399	2	Q9AEG2 enterobacte
9	760.5	37.7	409	16	Q8CWA9 escherichia
10	758	37.5	409	16	Q8XRL2 raietonia s
11	756	37.4	397	16	Q8ZBT7 salmonella
12	753	37.3	397	16	Q8ZRA6 salmonella
13	752.5	37.3	396	2	Q8VPA9 proteus mir
14	749.5	37.1	385	16	Q8X4W2 escherichia
15	745.5	36.9	385	16	Q8CVN7 escherichia
16	743.5	36.8	394	2	Q9F241 xanthomonas

17	741.5	36.7	404	16	Q8P1Q1 xanthomonas
18	734	36.4	382	16	Q8E8H1 shewanella
19	730.5	36.2	384	2	Q9WZ29 pseudomonas
20	727.5	36.0	408	16	Q8PQJ6 xanthomonas
21	726.5	36.0	395	16	Q8ZC88 yersinia pe
22	723	35.8	427	16	Q8UB14 agrobacteri
23	720	35.7	371	2	Q9KJC3 pseudomonas
24	719.5	35.6	407	16	Q9ZT02 rhizobium m
25	719	35.6	374	2	Q8GCB4 enterobacte
26	701.5	34.7	384	16	Q8CW42 escherichia
27	700.5	34.7	388	2	Q8RSM2 uncultured
28	699.5	34.6	387	16	Q51395 pseudomonas
29	697.5	34.5	373	16	Q8X7E1 escherichia
30	683	33.8	398	2	Q93K41 klebsiella
31	670	33.2	408	16	Q9PBP7 xylella fas
32	662	32.8	398	2	Q9RBY9 xanthomonas
33	655	32.4	385	16	Q8X4L0 escherichia
34	655	32.4	385	16	Q8CVL1 escherichia
35	654	32.4	382	2	Q9KWV5 pseudomonas
36	647.5	32.1	436	16	Q8P876 xanthomonas
37	647	32.0	396	2	Q93E20 actinobact
38	640	31.7	382	2	Q31099 pseudomonas
39	636	31.5	435	16	Q8P1N2 xanthomonas
40	631	31.3	391	2	Q93P05 pseudomonas
41	628.5	31.1	386	16	Q8ZBD3 yersinia pe
42	628	31.1	396	16	Q9RG60 pseudomonas
43	607.5	30.1	399	2	Q87935 burkholderi
44	606	30.0	385	16	Q8X704 raietonia s
45	602	29.8	384	2	Q68440 agrobacteri

ALIGNMENTS

RESULT 1

Q9J50 PRELIMINARY; PRT; 412 AA.

AC Q9J50; (TREMREL1, 15, Created)

DT 01-OCT-2000 (TREMREL1, 15, Last sequence update)

DT 01-OCT-2002 (TREMREL1, 22, Last annotation update)

DE Membrane fusion protein.

GN MTRC OR NMA1970.

OS Neisseria meningitidis (serogroup A).

OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;

OX Neisseriaceae; Neisseria.

OX NCBI_TaxID=5699;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Z2491 / Serogroup A / Serotype 4A;

RX MEDLINE=20222556; PubMed=10761919;

RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C., Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T., Davies R.M., Davis P., Devlin K., Felwell T., Hamlin N., Holtroyd S., Jagers K., Leather S., Moule S., Mungall K., Quail M.A., Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J., Whitehead S., Spratt B.G., Barrall B.G.,

RA "complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491."

RT Nature 404:502-506(2000).

RL EMBL; AL162757; CAB85190.1; -

DR InterPro; IPR006143; H1YP.

DR Pfam; PF00529; H1YP; 1.

KW Complete proteome.

SQ SSOURCE 412 AA; 42954 MW; SCF797BA370AA75D CRC64;

Query Match 98.7%; Score 1993; DB 16; Length 412;

Best Local Similarity 98.5%; Pred. No. 5,8e-102; Matches 406; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 MAFYAFKRAMRAALAAVALVSSCGKGGDAQGGPAGREAPAPVGVTVHPQVALT 60

DB 1 MAFYAFKRAMRAALAAVALVSSCGKGGDAQGGPAGREAPAPVGVTVHPQVALT 60

```
QY 61 VELPGRLESLRTADVRAVGGIIQKRLFOEGSYVRAQGPLYQIDSSYTEANLESARAQLA 120
Db 61 VELPGRLESLRTADVRAVGGIIQKRLFOEGSYVRAQGPLYQIDSSYTEANLESARAQLA 120
QY 121 TAQATLAKADADLARRYKPLVAEAVSRQEDAAVTAKRSABAGVKAQAQAIKSAGINLNR 180
Db 121 TAQATLAKADADLARRYKPLVAEAVSRQEDAAVTAKRSABAGVKAQAQAIKSAGINLNR 180
QY 121 TAQATLAKADADLARRYKPLVAEAVSRQEDAAVTAKRSABAGVKAQAQAIKSAGINLNR 180
Db 121 TAQATLAKADADLARRYKPLVAEAVSRQEDAAVTAKRSABAGVKAQAQAIKSAGINLNR 180
QY 181 SRITAPISGFGQSKVSEGTLLNAGDTTVLATTIQTNPMTVNTQSAEVMKLRQIAEG 240
Db 181 SRITAPISGFGQSKVSEGTLLNAGDTTVLATTIQTNPMTVNTQSAEVMKLRQIAEG 240
QY 181 SRITAPISGFGQSKVSEGTLLNAGDTTVLATTIQTNPMTVNTQSAEVMKLRQIAEG 240
Db 181 SRITAPISGFGQSKVSEGTLLNAGDTTVLATTIQTNPMTVNTQSAEVMKLRQIAEG 240
QY 241 KLLAADGVIAVGIKFDGTVYPEKGRLLPADPVNVESTGOITLRAVPNDONILMPGLYV 300
Db 241 KLLAADGVIAVGIKFDGTVYPEKGRLLPADPVNVESTGOITLRAVPNDONILMPGLYV 300
QY 241 KLLAADGVIAVGIKFDGTVYPEKGRLLPADPVNVESTGOITLRAVPNDONILMPGLYV 300
Db 241 KLLAADGVIAVGIKFDGTVYPEKGRLLPADPVNVESTGOITLRAVPNDONILMPGLYV 300
QY 301 RVLMDQVAVDNAFVVPQAAVTRGAKDTVMIVNAQGMPEPREVTVAQOQGTNMIYVTSGLKD 360
Db 301 RVLMDQVAVDNAFVVPQAAVTRGAKDTVMIVNAQGMPEPREVTVAQOQGTNMIYVTSGLKD 360
QY 361 GDKVVEGISIAGITGAKKVTPEKMASSENOAAAPQSGVOTASAKTASEAE 412
Db 361 GDKVVEGISIAGITGAKKVTPEKMASSENOAAAPQSGVOTASAKTASEAE 412

RESULT 2
ID Q9JY66 PRELIMINARY; PRT; 412 AA.
AC Q9JY66;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Membrane fusion protein.
GN NMB1716.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
NC NCB1_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tectelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Hatt D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Macon T., Ciecko A., Parksey D.S., Blair E., Clifton H., Clark E.B.,
RA Cotton M.D., Ullrich T.R., Kouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Masiann V., Pizzo M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.,
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RL Science 287:1809-1815 (2000).
DR EMBL; AE002521; AAF42063.1; -.
DR TIGR; NMB1716; -.
DR InterPro; IPR006143; HLYD.
DR Pfam; PF00529; HLYD; 1.
KW Complete proteome.
SQ SEQUENCE 412 AA; 42795 MW; 0B7A0DB0E3113E4E CRC64;

Query Match 98.5%; Score 1989; DB 16; Length 412;
Best Local Similarity 98.5%; Pred. No. 9.7e-102;
Matches 406; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 121 TAQATLAKADADLARRYKPLVAEAVSRQEDAAVTAKRSABAGVKAQAQAIKSAGINLNR 180
Db 121 TAQATLAKADADLARRYKPLVAEAVSRQEDAAVTAKRSABAGVKAQAQAIKSAGINLNR 180
QY 181 SRITAPISGFGQSKVSEGTLLNAGDTTVLATTIQTNPMTVNTQSAEVMKLRQIAEG 240
Db 181 SRITAPISGFGQSKVSEGTLLNAGDTTVLATTIQTNPMTVNTQSAEVMKLRQIAEG 240
QY 181 SRITAPISGFGQSKVSEGTLLNAGDTTVLATTIQTNPMTVNTQSAEVMKLRQIAEG 240
Db 181 SRITAPISGFGQSKVSEGTLLNAGDTTVLATTIQTNPMTVNTQSAEVMKLRQIAEG 240
QY 241 KLLAADGVIAVGIKFDGTVYPEKGRLLPADPVNVESTGOITLRAVPNDONILMPGLYV 300
Db 241 KLLAADGVIAVGIKFDGTVYPEKGRLLPADPVNVESTGOITLRAVPNDONILMPGLYV 300
QY 241 KLLAADGVIAVGIKFDGTVYPEKGRLLPADPVNVESTGOITLRAVPNDONILMPGLYV 300
Db 241 KLLAADGVIAVGIKFDGTVYPEKGRLLPADPVNVESTGOITLRAVPNDONILMPGLYV 300
QY 301 RVLMDQVAVDNAFVVPQAAVTRGAKDTVMIVNAQGMPEPREVTVAQOQGTNMIYVTSGLKD 360
Db 301 RVLMDQVAVDNAFVVPQAAVTRGAKDTVMIVNAQGMPEPREVTVAQOQGTNMIYVTSGLKD 360
QY 361 GDKVVEGISIAGITGAKKVTPEKMASSENOAAAPQSGVOTASAKTASEAE 412
Db 361 GDKVVEGISIAGITGAKKVTPEKMASSENOAAAPQSGVOTASAKTASEAE 412

RESULT 3
ID Q51007 PRELIMINARY; PRT; 271 AA.
AC Q51007;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE MtrC protein (Fragment).
GN MTRC.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
NC NCB1_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CH95;
RX MEDLINE=94254732; PubMed=8196548;
RA Pan W., Spratt B.G.;
RT "Regulation of the permeability of the gonococcal cell envelope by the
RT mtr system."
RL Mol. Microbiol. 11:769-775 (1994).
DR EMBL; Z25796; CAAB1046.1; -.
DR InterPro; IPR006143; HLYD.
DR Pfam; PF00529; HLYD; 1.
KW NON_TER
SQ SEQUENCE 271 AA; 28145 MW; 1D0ADD1F335B39C5 CRC64;

Query Match 63.3%; Score 1278; DB 2; Length 271;
Best Local Similarity 96.7%; Pred. No. 6.6e-63;
Matches 262; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
```

RESULT 4
Q8Y3G9 PRELIMINARY; PRT; 398 AA.
AC Q8Y3G9;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DE Probable acriflavin resistance lipoprotein A precursor.
GN ACRA OR RSC0011 OR RSO1833.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Ralstoniaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GMI1000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brothier P., Camus J.C., Catolico L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cunac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Sautin W., Schlex T.,
RA Signier P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weissbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum";
RL Nature 415:497-502(2002).
DR EMBL; AL646057; CADI3539.1; -
DR InterPro; IPR006143; HLYD.
DR Pfam; PF00529; HLYD; 1.
KW Complete proteome.
SQ SEQUENCE 398 AA; 41234 MW; F5E448BA0583F27C CRC64;
Query Match 41.4%; Score 836; DB 16; Length 398;
Best Local Similarity 47.0%; Pred. No. 2,2e-38;
Matches 188; Conservative 71; Mismatches 127; Indels 14; Gaps 8;
QY 14 LAAVALVSSCGGGAAGCGGAPAGREAPAVGVTVH--PQVATVTELPGRLESRTA 73
DB 10 LAAAGLTVVLAACGNKQACG--PGGMPTE-VGVTVVQPHSVGLTTELPGRLEATRV 66
QY 74 DVRAOVGGIIOKRIFOGSSVYRAGOPLYOIDSTYEANLESARQATLTAQTLAKADL 133
DB 67 QVRARVAGIVAKRTYQBSGDVKNADVFRIDPAQYASLSAKQALAAEXTQTOAQOLKA 126
QY 134 ARYEPVLAEEAVSHOEYDAATVAKRSAGVKAQAQAIKSGININRSRTAPISGFIGQ 193
DB 127 ERYKPELVATNAISQDDDDAAAKQATADGARAARAVETKALNGATVTSISGAAGL 186
QY 194 SKVSEGTLLNAG--DTVLTATIRQTNPMVYVNTQASAEVMKLRQIAEGKLAA--DGVIAV 251
DB 187 AQTVEGALVGGSDATILLATVQIDPIYLFPTQSTEWRLQELAKGLAAGDTAAKV 246
QY 252 GIKRDDGVYPERKRLFPADPVNVESTGQITLRAAVPNDQNIIMPGLYRVLMQVAVDN 311
DB 247 TLVTEDEVRVAAGTQKLYFSDLTVDTTGSITLRAIFPAERTLLPGMYVRAARLEQAVDQ 306
QY 312 AFVVPQAAVTRGAKD--TWIYVNAOGGMEPREVTVAOOGTNIWTSGLKDGKVVVEGIS 370
DB 307 AITVPPQAAVGRGADGASVMIVDAEGKVAAPRVQADRAVGIDMIVSSGLKAGDKYIVDGLQ 366
QY 371 IAGITGAKKVTPEKMASSENOAAAPQSGVQTASEAKTASE 410
DB 367 --KVRPGAPVKPVAM-----QAGGAQ--GCGAASAPAKQ 398

GN ACRA OR STM3390.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=L72 / SCS1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McLelland M., Sanderson K.E., Spierh J., Clifton S.W., Latreille P.,
RA Courtney L., Portwollik S., Ali U., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan B., Sun H., Florea L., Miller W., Stoneking T., Nhan W.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2";
RL Nature 413:852-856(2001).
DR EMBL; AE008856; ALU2259.1; -
DR InterPro; IPR006143; HLYD.
DR Pfam; PF00529; HLYD; 1.
KW Complete proteome.
SQ SEQUENCE 385 AA; 40963 MW; 230969D5908572B CRC64;
Query Match 38.5%; Score 777; DB 16; Length 385;
Best Local Similarity 42.5%; Pred. No. 3.6e-35;
Matches 170; Conservative 71; Mismatches 121; Indels 38; Gaps 5;
QY 19 ALVLSGCGKGGDA--AGCGAPAGREAPAVGVTVH--PQVATVTELPGRLESRTAD 74
DB 18 AALLAGCNDGSDTOAHAGBEPD-----VTVHVEVETPLAVTTELPGRISAFRIAE 66
QY 75 VRAOVGGIIOKRIFOGSSVYRAGOPLYOIDSTYEANLESARQATLTAQTLAKADL 134
DB 67 VRPQVSGIVAKRNTFESDVBAQSLQIDPATYQADYDAKGLASBAALAHITLVK 126
QY 135 RYKPELVAAEAVSHOEYDAATVAKRSAGVKAQAQAIKSGININRSRTAPISGFIGQ 194
DB 127 RYVPLVGTGYISQOEYQADADARQADAAVAAKAAVESARINLAVTKVTSISGRIGKS 186
QY 195 KVESEGTLLNAGDTVLTATIRQTNPMVYVNTQASAEVMKLRQIAEGKLAAAGVIAVG 254
DB 187 NVTEGALVNTGQSTELTAVQIDPIYLVDTQSSNDFRLKQSVQGLHSDASSTVQLV 246
QY 255 FDDGTVYPERKRLFPADPVNVESTGQITLRAAVPNDQNIIMPGLYRVLMQVAVD 314
DB 247 MENGOVYPIGTIQFSVTVYDESGITTLRAVPNPHSHLPGMFVYARIDEGQPAAIL 306
QY 315 VPQAAVTRGAKD--TWIYVNAOGGMEPREVTVAOOGTNIWTSGLKDGKVVVEGIS 372
DB 307 VPQGVTRTRGDAWVNVVNDKQVEARNVVAAQALIDKWLISGLKPGKVIIVSGL--- 363
QY 373 GITGAKKVTPEKMASSENOAAAPQSGVQTASEAKTASE 412
DB 364 -----QARPGVQVKAITDPAAKTAQ 385

RESULT 6
Q9AA03 PRELIMINARY; PRT; 388 AA.
AC Q9AA03;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE HlyD family secretion protein.
GN CC0808.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
OC Caulobacteraceae; Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;

RA Nieman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
 RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
 RA Porocka J., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
 RA Deboy R.T., Dodson R.J., Dirkin A.S., Gwinn M.L., Hafe D.H.,
 RA Kolonay J.F., Smit J., Craven M.B., Khoult H., Shetty J., Berry K.,
 RA Osterberg T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.,
 RA "Complete genome sequence of *Caulobacter crescentus*,"
 RT Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
 RL EMBL; AE005757; AAK22793.1; -.
 DR TIGR; CC0808; -.
 DR InterPro; IPR006143; HlyD.
 DR Pfam; PF00529; HlyD; 1.
 KW Complete proteome.
 SQ SEQUENCE 388 AA; 39483 MW; 409343CF69052DAD CRC64;

Query Match 38.1%; Score 770; DB 16; Length 388;
 Best Local Similarity 46.2%; Pred. No. 8.9e-35;
 Matches 180; Conservative 57; Mismatches 131; Indels 22; Gaps 9;

OY 16 AAVLVVSSCG-KGSDAAGGPPAGREAPVGVVTHPQVTLTVLPRLSLRTAD 74
 DB 15 AAIALLTASCCQKPG---GEMGEMGGPTE-VGVIVAQSOSVGLSTELAGRTSVYVSE 69
 OY 75 VRAOVGIIORLFOEGSVYRAGOPLYOIDSSTYANLESARAPL-ATAQATLAKAD 132
 DB 70 VRPOVSGVIOKRLPEBGAVRAGOPLYOIDPATYAAANSAALAAQAOATLAKAD 129
 OY 133 LARYKPLVAEAVSRQEYDAVTAKRSAEAGVKAQAAIKSAGINLNSRTTAPISGIG 192
 DB 130 --RYKALVETGAVSRQDNDAAALQTPAAVGVKALDSARINLVNARYTAPISGRIG 187
 OY 193 OSKSYSEGLTAAAGDTTLATIRQTNPMYVNTQSAEFMKLRQIAEKLAAQVAVG 252
 DB 188 KSSVYAGALVLANQATLALATVODLSKVYVDLTQTSABELKLOAQPSAGK-VGRSGSAQVT 246
 OY 253 IKFDGTVYPERGRLLFADPVVNNESTQITLRAAVPNDQNIIMPGLYRVLMQVAVDNA 312
 DB 247 LKLEBGSYVPIRGRLEFSDITVDPTGAVGLRAVDNKKGVLLPMTYRAVAVLSQVAVASG 306
 OY 313 FVVPQAVTRAKD--TWIYVNAQGMREPVYVNAQOGTNMIVTSGKDGKVVESIG 370
 DB 307 ILIIPOTAVNRDPKGGATVMLVGAK-GPEPRPVTLIGQTVGDKWLVTSGLNAGDKVIVEGL- 364
 OY 371 IAGITGAKKVTPEKWEASSENQAAAPQSGVQ 400
 DB 365 -----MKVRPGADIKAVPAGAAPAAQ 387

RESULT 7

O8P7C8 PRELIMINARY; PRT; 399 AA.
 AC O8P7C8;
 DT 01-OCT-2002 (TREMBlrel. 22, Created)
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Multidrug resistance protein.
 GN MEYA OR KCC2693.
 OS *Xanthomonas campestris* (pv. *campestris*).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xanthomonas.
 OX NCBI_TaxID=340;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 33913 / NCEPB 528;
 RX MEDLINE=22022145; PubMed=12024217;
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
 RA Oussiglo R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
 RA Alves L.M.C., do Amaral A.M., Betrolini M.C., Camargo L.E.A.,
 RA Camarotte G., Cannavan F., Cardoso J., Chambeiro F., Clapina L.P.,
 RA Ciccarelli R.M.B., Coutinho L.D., Cursino-Santos J.R., El-Dorri H.,
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
 RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,

RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
 RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
 RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
 RA Pereira L.A., Rossi A., Sana J.A.D., Silva C., de Souza R.F.,
 RA Spindola H.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
 RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
 RA Terubal J.C., Kitajima J.P.;
 RT "Comparison of the genomes of two *Xanthomonas* pathogens with differing
 RT host specificities";
 RL Nature 417:459-463(2002).
 RL EMBL; AE012380; AAM41955.1; -.
 DR InterPro; IPR006143; HlyD.
 DR Pfam; PF00529; HlyD; 1.
 KW Complete proteome.
 SQ SEQUENCE 399 AA; 41511 MW; 59C1197DE60E5D2D CRC64;

Query Match 37.9%; Score 765; DB 16; Length 399;
 Best Local Similarity 43.3%; Pred. No. 1.7e-34;
 Matches 174; Conservative 67; Mismatches 139; Indels 22; Gaps 6;

OY 8 AMRAALAAVALVSSCGK--GGDAAGGPPAGREAPVGVVTHPQVTLTVLPRLSLRTAD 65
 DB 6 SLRLPLAAVASVLSACGSPGGPPQG-----TPVGVITVRPOPVTTLTQLPG 57
 OY 66 RLSELTADYRAOVGIIORLFOEGSVYRAGOPLYOIDSSTYANLESARAPLATAQAT 125
 DB 58 RTVYVLAEBRPQGVIVOTRPFEGSDVAKAGOTLYOIDPATYAAANSAALAAQAOATLAKAD 117
 OY 126 LAKADADARYKPLVAEAVSRQEYDAVTAKRSAEAGVKAQAAIKSAGINLNSRTTAPISGIG 185
 DB 118 LRTARLADRYKELVOIKALISQEGDDTAATLGAQEAADVAAKASVETARINLAFAMDA 177
 OY 186 PISGFIOQSKVSEGLTAAAGDTTLATIRQTNPMYVNTQSAEFMKLRQIAEKLAAQVAVG 244
 DB 178 PISGRIRSSVTPALVTAQATLALTIQJLDPIYDVTPSAVVLKKAAMAGDLARA 237
 OY 245 ADGVIAVGIFKFDGTVYPERGRLLFADPVVNNESTQITLRAAVPNDQNIIMPGLYRVLM 304
 DB 238 GDGAQVSVLVEDSDITPLOGRLAFSDVTVQNTGISLRAVFPNPADLLPMTYRAVAVLSQ 297
 OY 305 DQVAVDNAFVVPQAVTR--GAKDTVMVNAQGMREPVYVNAQOGTNMIVTSGKDG 362
 DB 298 QEGVKAQGVLPQAVTRNAGKPTAFVGAADNKLQRLVETBRAVDQWLVRSGLKTGD 357
 OY 363 KVVVEGISIAGITGAKKVTPEK-----ASSNQAAAPQS 397
 DB 358 QLVVDGLSRARDGVQVKTVP--WQPTSASVAGSPSAPAAPRA 397

RESULT 8

O9AEG2 PRELIMINARY; PRT; 399 AA.
 AC O9AEG2;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Acra protein precursor.
 GN ACRA.
 OS Enterobacter aerogenes (aerobacter aerogenes).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Enterobacter.
 OX NCBI_TaxID=548;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BW16627;
 RX Pradel B., Pages J.M.;
 RA "The *Acra*/AcraB/olc efflux pump participates in multidrug resistance
 RT in *Enterobacter aerogenes*,"
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ306389; CAC35724.1; -.
 DR InterPro; IPR006143; HlyD.
 DR Pfam; PF00529; HlyD; 1.

KM Signal. 1 24 POTENTIAL.
 FT SIGNAL 25 399 ACRA.
 FT CHAIN 25 399 ACRA.
 SQ SEQUENCE 399 AA; 42443 MM; AC49F8E3870B6E78 CRC64;

Query Match 37.8%; Score 763.5; DB 2; Length 399;
 Best Local Similarity 42.0%; Pred. No. 2.1e-34;
 Matches 169; Conservative 80; Mismatches 134; Indels 19; Gaps 5;

QY 11 AALAAALVALVSSCGGGAAGGAGPAGREAPAVVTVVHPTQVATLVELGRLLESL 70
 DB 11 AVTLMSSSLATGCG-DKPKAQQGAGQ-----QMPVGIITLKSAPIQITTELPRTINAY 63
 QY 71 RTADVRAQVGGIIQKRLFOEGSYVRAGOPLYQIDSTYEANLESARQATATQATLAKAD 130
 DB 64 RVAEVRPQVSGIILKRNFTGESSDIAQGVSLYQIDPATYQAYESAKGDLAKQAQAANIQA 123
 QY 131 ADLARYKPLVAAEAVSRQYDAVATKRSAAAGYKAAQAALIKSAGININRSRTAPISGF 190
 DB 124 LTVRRYQKLGTYKYSIQOEYDAAVADAQOSNAAVAAVETARINILATYKVTSPISGR 183
 QY 191 IGSQKSEGTLLNAGDTTVALTITQTPMNVNTQSSSEWKLRRQIAEGKLLAADGVIA 250
 DB 184 IGKSAVTEGALVQNGOSTALATVQQLDPIYDVQSSNDLRLKQELANGTLKQENGKAK 243
 QY 251 VGIKFDGTIVPEKGRLLFADPVVNESTGQITLRAAVPNQNIIMPGLYVRLMDQVAVD 310
 DB 244 VELVTNGLKAKYPOGBTLEFSDVTVDTGTSITLRAIFPNPDHITLLPGMFARAREBGINP 303
 QY 311 NAFVVPQQAATVRGAKD--TWIIVNAQGGMEPREVTVAQOOGTNNIIVTSGKDGKRVVEG 368
 DB 304 DALIVPQOGVTRTPRGDPAVLVGCADDKVETRPVIAQALGDKVLVTEGLKAGDRVIVAG 363
 QY 369 ISAGITGAKKVTPEKMASSENOAAAPQSGVQTAASEAKTASE 410
 DB 364 L-----QKVRGQVQKAEVYADNQQ--QTAGNNAQSE 395

RESULT 9

08CWA9 PRELIMINARY; PRT; 409 AA.
 AC 08CWA9;
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DE Acriflavine resistance protein A precursor.
 GN ACRA OR C0581.
 OS Escherichia coli O6.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=217992;
 RN [1]
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum."
 RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
 RA Mayhew G.P., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RA Mobley H.L.T., Domeneberg M.S., Blattner F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 of uropathogenic Escherichia coli."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
 DR EMBL; AEO16756; AAN79059.1; -.
 KW Complete proteome.
 SQ SEQUENCE 409 AA; 43521 MM; 4234179F03G60197 CRC64;

Query Match 37.7%; Score 760.5; DB 16; Length 409;
 Best Local Similarity 42.2%; Pred. No. 3.2e-34;
 Matches 166; Conservative 76; Mismatches 129; Indels 25; Gaps 6;

QY 11 AALAAALVALVSSCGGGAAGGAGPAGREAPAVVTVVHPTQVATLVELGRLLESL 70
 DB 23 AVTLMSSSLATGCG-DKPKAQQGAGQ-----QMPVGIITLKSAPIQITTELPRTINAY 75

QY 71 RTADVRAQVGGIIQKRLFOEGSYVRAGOPLYQIDSTYEANLESARQATATQATLAKAD 130
 DB 76 RVAEVRPQVSGIILKRNFTGESSDIAQGVSLYQIDPATYQAYESAKGDLAKQAQAANIQA 135
 QY 131 ADLARYKPLVAAEAVSRQYDAVATKRSAAAGYKAAQAALIKSAGININRSRTAPISGF 190
 DB 136 LTVRRYQKLGTYKYSIQOEYDAAVADAQOSNAAVAAVETARINILATYKVTSPISGR 195
 QY 191 IGSQKSEGTLLNAGDTTVALTITQTPMNVNTQSSSEWKLRRQIAEGKLLAADGVIA 250
 DB 196 IGKSAVTEGALVQNGOSTALATVQQLDPIYDVQSSNDLRLKQELANGTLKQENGKAK 255
 QY 251 VGIKFDGTIVPEKGRLLFADPVVNESTGQITLRAAVPNQNIIMPGLYVRLMDQVAVD 310
 DB 256 VSLTSPGKIFPOGBTLEFSDVTVDTGTSITLRAIFPNPDHITLLPGMFARAREBGINP 315
 QY 311 NAFVVPQQAATVRGAKD--TWIIVNAQGGMEPREVTVAQOOGTNNIIVTSGKDGKRVVEG 368
 DB 316 NAFVVPQQAATVRGAKD--TWIIVNAQGGMEPREVTVAQOOGTNNIIVTSGKDGKRVVEG 375
 QY 369 ISAGITGAKKVTPEKMASSENOAAAPQSGVQTAASEAKTASE 400
 DB 376 L-----QKVRGQVQKAEVYADNQQAA--SGAQ 403

RESULT 10

08XRL2 PRELIMINARY; PRT; 409 AA.
 ID 08XRL2;
 AC 08XRL2;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
 DE Probable drug efflux lipoprotein.
 GN RSP0819 OR RS0188.
 OS Ralstonia solanacearum (Pseudomonas solanacearum).
 OG Plasmid megaplasmid.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Ralstoniaceae; Ralstonia.
 OX NCBI_TaxID=305;
 RN [1]
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum."
 RA Ariat M., Billaud A., Broctier P., Camus J.C., Catalicco B.,
 RA Chander M., Choisme N., Claudel-Renard C., Cunac S., Demange N.,
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
 RA Sigulier P., Thebaud P., Whalen M., Wincker P., Levy M.,
 RA Weisenbach J., Boucher C.A.;
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum."
 RL Nature 415:497-502(2002).
 DR EMBL; AL646081; CAD17970.1; -.
 DR InterPro; IPR006143; HLYD.
 DR Pfam; PF00529; HLYD; 1.
 KW Plasmid; Complete proteome.
 SQ SEQUENCE 409 AA; 42873 MM; 97003B2B7F95401B CRC64;

Query Match 37.5%; Score 758; DB 16; Length 409;
 Best Local Similarity 44.4%; Pred. No. 4.4e-34;
 Matches 182; Conservative 68; Mismatches 136; Indels 24; Gaps 9;

QY 12 AALAAALVALVSSCGGGAAGGAGPAGREAPAVVTVVHPTQVATLVELGRLLESL 68
 DB 13 AALAA--ASLLAAGC-----PPGPPPEAGTTPVGVVTVVQPORVTLDTLELPGRTV 61
 QY 69 SLRTADVRAQVGGIIQKRLFOEGSYVRAGOPLYQIDSTYEANLESARQATATQATLAK 128
 DB 62 PFLVADVQVQNGIITAKKFRBESDVKAGALYQIDPATYQAYDSNVAAALAKQAANLKT 121
 QY 129 ADLARYKPLVAAEAVSRQYDAVATKRSAAAGYKAAQAALIKSAGININRSRTAPIS 188
 DB 122 TRUKARHKEIVAIQAVSRQYDDDAALAAQEGADVAARAVETSRINILAVARVAPIS 181

Oy		189	GFIGGSXKVS	EGTLNMAAGPTVYATIRQNNPMWVNTQSGASVFMKR.RRQIAEKKLAADGV	248			
Dd		182	GRIGSSVT	PGLVLANOTSTISLTIQQDDPIYDVDTGPSAALLRLURQANARDDLQSKSVAN	241			
Oy		249	IA-VAIKFDDGTV	VPEKGRLLEPADVVNESTGOITLRAPVPNDQINLMPLGYVRVLMDQV	307			
Dd		242	AATVRLLEDGS	AVYLEGLEFSVDVYVQNTSSTVLRANFRPSPAILLPQMYYRAVLPEBG	301			
Oy		308	AYDNAFVVPQA	AVTGA--KDTVMIVNAOGMEPREBVTVAAQOQGTMVIYTSGLKDGKVY	365			
Dd		302	VDEAEALLVPOQA	VARDSGTCKPRPAVYVGNDRKLQRRTLETRETFRVGGQMVRSGLRICDOLV	361			
Oy		366	VEGISIAGITGAK	-RYTPKEW---ASSENQAAAPQSOGVOTASEATASRA	411			
Dd		362	VEGLPRA-VPGAEVKTTP-	-WTGKTATISNPAAAPAAVATAVGORTVAA	408			
<hr/>								
	RESULT	11						
ID	0828T7		PRELIMINARY;	PRT;	397 AA.			
AC	0828T7							
DT	01-MAR-2002	(TREMBLrel. 20, Created)						
DT	01-MAR-2002	(TREMBLrel. 20, Last sequence update)						
DT	01-OCT-2002	(TREMBLrel. 22, last annotation update)						
DE	Acridinavin resistance protein A.							
CN	STR0520.							
OS	Salmonella typhi.							
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;							
CC	Enterobacteriaceae; Salmonella.							
Ox	NCEI_TaxID=601;							
RN	[1]							
RP	SEQUENCE FROM N.A.							
RC	STRAIN=CT18;							
EX	MEDLINE=21534947; PubMed=11677608;							
RA	Churchill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,							
RA	Baker S., Bangham A.L., Bentley S.D., Holden M.T.G., Sebahia M.,							
RA	Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,							
RA	Feltwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagals K.,							
RA	Krogh A., Larsen T.S., Leachner S., Moule S., O'Gaora P., Parry C.,							
RA	Quell M., Rutherford K., Simmonds M., Skelton J., Stevens K.,							
RT	"Complete genome sequence of a multiple drug resistant Salmonella							
RL	enterica serovar Typhi CT18.";							
RT	Nature 413:848-852(2001).							
DR	EMBL; AL627267; CAD04961.1; --							
DR	InferPro; IPR006143; HlyD.							
DR	Pfam; PF00529; HlyD; 1.							
KV	Complete proteome.							
SQ	SEQUENCE 397 AA; 42221 MW; F867D42FD8B9EE8D CRC64;							
<hr/>								
	Query Match	37.4%;	Score 756;	DB 16;	Length 397;			
	Best Local Similarity	41.3%;	Pred.No. 5.4e-34;					
	Matches 166;	Conservative	78;	Mismatches 128;	Indels 30; Gaps 5			
Oy		11	AAALAANAALVLS	SCGKGDAAGQGPARGREKRAPVNGVTHNPQTVALTYVLPRLSL	70			
Dd		11	AVLTMLSSSLATGCDKDD	QQGQ-----QMFEVGVTILKTERLDITTELPGRYAV	63			
Oy		71	RTADVRAOVGGII	IQRLFOEGSYVRAGOPLYVIDSITYEANLESARAQDLATQAUTLAD	130			
Dd		64	RIAEVRPVQSGIIL	TKRNFEVSGSDIEAGVSLYVIDPATYQATYDASKGDLAKQAAMANI	123			
Oy		131	ADLARPKRLVA	EANSROETPDAAVTAKSAEAGVYAQAQAALYSAGINLMRSKITPISGF	190			
Dd		124	LTVKRYQQLTGTOY	ISKOEYODALADAQAQATAVVAANAARETAIINALYTKVTSPISGR	183			
Oy		191	IGOSVSEGTLL	INADDTVLIATIRQTNPNPYVNVVOTOSASFVMKLRRQIAEGKLLAADGYIA	250			
Dd		184	IGKSSVTGALV	ONQASALATVQDDLPDYVDVDTOSSNDFLRKELANGSLKQENGAKR	243			
Oy		251	VGIRKDDGTVV	PEKGRLLFPADVNVESGTQITLRAPVPNDQINLMPLGYVRVLMDQAVD	310			

[illegible]

DB 364 L-----QKVRPGAQYKVGQETITADNKGQAASGDPQAQPRS 397

RESULT 13

Q8VPA9 PRELIMINARY; PRT; 396 AA.

AC Q8VPA9

DT 01-MAR-2002 (Tremblrel. 20, Created)

DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)

DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)

DE Membrane fusion protein Acra.

OS Proteus mirabilis.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Proteus.

NCBI_TaxID=584;

SEQUENCE FROM N.A.

RA Visalli M.A., Murphy E., Projan S.J., Bradford P.A.;

RT "Acr multidrug efflux homolog is associated with reduced levels of

RT susceptibility to GAR-936 in Proteus mirabilis."

RL Submitted (NOV-2001) to the EMBL/Genbank/DBJ databases.

DR EMBL; AY061647; AAL32125.1; -

DR InterPro; IPR006143; HlyD.

DR Pfam; PF00529; HlyD; 1.

SO SEQUENCE 396 AA; 42256 MW; 261F3865A4746A23 CRC64;

Query Match 37.3%; Score 752.5; DB 2; Length 396;

Best Local Similarity 42.6%; Pred. No. 8.4e-34;

Matches 168; Conservative 71; Mismatches 138; Indels 17; Gaps 6;

QY 12 AALAAVALVSSCGKGDAAAGGQAPAGREAPVGVTVHPQTALTVLPGRLESIR 71

DB 11 ALLVLSGSLVAVAGCGDKNQSA-GGPP-----PAPAVGVTLDAKPLRTTDLPGRTSAYR 64

QY 72 TADVRAQVGGIIQKRLFOGSSYVRAGOPLYOIDSSTYANLBSARAQATQAATLAKDA 131

DB 65 IAEVRPQVGGIILKKNYSSIVEKSLYQIDPITPATNSAQADLAKKAAAEIARL 124

QY 132 DLAVYKPLVAEAVSRQEDAAVTAKRSAEAGVKAQAQAIKSAGINLNSRTAPISGFI 191

DB 125 TVERYKPLGLTGNVYSKQDFRTISQYQAVAAVKAELVTVNKNLETKVTAPISGRS 184

QY 192 GQSKVSEGLTINAGDTVLATIRQTNPMYVNVTOSSASEVMKLRQIAEGKLLAADGVIAV 251

DB 185 GKSTVTGEGALVAPGQVALLTVQQLDPIYVDVTOSSSEYDLKLNKEIESGIIROEGKRVV 244

QY 252 GIKEDDGVYPEKGRLLFADPVVNESTQITLRAAVPNDQNIILPGLYVRVLMQVAVDN 311

DB 245 HLTITNSQSYAQKHLEFSDVTVDESTSITMRAIVPMPKGBELLPGMFVTRKLENGIRON 304

QY 312 AFVVPQQAATRGAKD--TVMIVNAQGGMEPREVTVAAQOQGTNMTVTSGLKXGDKVYVGEI 369

DB 305 AILPQQAATVTPRGGAATTVVNAKDNVVEVRTIEVSQAVGNKMLVNSGVQGDVIVSGL 364

QY 370 SIAGITGAKKVTPEK-----WASSENQAQA--PQ 396

DB 365 QKA--QPEMKVTPQEBNLDATASTEKSEPAQDPQ 396

RESULT 14

Q8X4W2 PRELIMINARY; PRT; 385 AA.

AC Q8X4W2

DT 01-MAR-2002 (Tremblrel. 20, Created)

DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)

DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)

DE Transmembrane protein affects septum formation and cell membrane

DE permeability.

GN ACRE OR Z4625 OR ECG4137.

OS Escherichia coli O157:H7.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

NCBI_TaxID=83334;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / EDL933 / ATCC 700927;

RX MEDLINE=21074935; PubMed=11206551;

RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,

RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,

RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,

RA Grotbeck E.J., Davis N.W., Lim A., Dimailanta E.T., Potamoukis K.,

RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,

RA Welch R.A., Blattner F.R.;

RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7,"

RL Nature 409:529-533(2001).

SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / RIMD 0509952;

RX MEDLINE=21156231; PubMed=11258796;

RA Haysheh T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,

RA Han C.-G., Ohtsuda E., Nakayama K., Murata T., Tanaka M., Tobe T.,

RA Iida T., Takami H., Honda T., Sasekawa C., Ogawara N., Yasunaga T.,

RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;

RT "Complete genome sequence of enterohaemorrhagic Escherichia coli

RT O157:H7 and genomic comparison with a laboratory strain K-12.";

RL DNA Res. 8:11-22(2001)

DR EMBL; AE005354; AAG58393.1; -

DR EMBL; AP002564; BAB37560.1; -

DR InterPro; IPR006143; HlyD.

DR Pfam; PF00529; HlyD; 1.

KW Complete proteome.

SO SEQUENCE 385 AA; 41345 MW; 9ABF02C84AF2F46B CRC64;

Query Match 37.1%; Score 749.5; DB 16; Length 385;

Best Local Similarity 42.1%; Pred. No. 1.2e-33;

Matches 168; Conservative 74; Mismatches 122; Indels 35; Gaps 6;

QY 17 AVALVSSCGKGPAAAGGQAPAGREAPVGVTVH-PQTVALTV--ELPGRLESIRTA 73

DB 17 SAALIAQCNDEGEKKAHVGEPO-----TVHIVKTAPELVKTELPGRTAAYRIA 65

QY 74 DVRAQVGGIIQKRLFOGSSYVRAGOPLYOIDSSTYANLBSARAQATQAATLAKADL 133

DB 66 EVRPQVGGIIVNRPFTGSDVQAGOSLYOIDPATYQAVDAKELASEAAAIHLITV 125

QY 134 ARYKPLVAEAVSRQEDAAVTAKRSAEAGVKAQAQAIKSAGINLNSRTAPISGFIQ 193

DB 126 KRYVPLVGTXYISQGEVDQALADARQDAVATAKAVESARINILATKYTAPISGRGX 185

QY 194 SKVSEGLTINAGDTVLATIRQTNPMYVNVTOSSASEVMKLRQIAEGKLLAADGVIAVG 253

DB 186 STVTGEGALVTVNGQTTTELATVQQLDPIYVDVTOSSSEYDLKLNKEIESGIIROEGKRVV 244

QY 254 KFDGTVYPEKGRLLFADPVVNESTQITLRAAVPNDQNIILPGLYVRVLMQVAVDNF 313

DB 246 VMENGQTYPLKGLTQFSDVTVDESTSITLRAVPNPQHTLLPGMFVTRKLENGIRON 305

QY 314 VVPQQAATRGAKD--TVMIVNAQGGMEPREVTVAAQOQGTNMTVTSGLKXGDKVYVGEI 371

DB 306 LIRPQQAATVTPRGGAATTVVNAKDNVVEVRTIEVSQAVGNKMLVNSGVQGDVIVSGL 363

QY 372 SIAGITGAKKVTPEK-----WASSENQAQA--PQ 396

DB 364 -----QKARPE-----QVKAATDTDPADTAK 385

RESULT 15

Q8CVN7 PRELIMINARY; PRT; 385 AA.

AC Q8CVN7

DT 01-MAR-2003 (Tremblrel. 23, Created)

DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)

DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)

DE Acridiflavine resistance protein B precursor.

GN ACRE OR C4031.

OS Escherichia coli O6.

OC Bacteri: Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_taxid=217992;
RN [1]
RP SEQUENCE FROM N.A.
RP STRAIN=O6:H1 / CFT073 / ATCC 700928;
RC MEDLINE=22368234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Raeko D., Buckles E.L., Liu S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Wolfe H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
KM EMBL: AE016767; AAN82471.1; --
DN Complete proteome.
SQ SEQUENCE 385 AA; 41289 MW; 5BB68F95B67C496 CRC64;

Query Match	36.9%	Score	745.5	DB	16	Length	385
Best Local Similarity	41.9%	Pred. No.	2e-33				
Matches	167	Conservative	75	Mismatches	122	Indels	35
						Gaps	6

```

0Y      17  AVALUWSCGKGDAAGGOPAGREAPAVVGWTVH--PQVATLV--ELPGRLESRTA 73
      18  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0b      17  SAALTAGCNDCKEEAAHVGEFO-----VYHIVKTPLEVKTELPRTYAIRIA 65
      18  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y      74  DVRAOVGGIIOKRLFOEGSYVPAPOPLYOIDSSSTEAMLESARLOATAQATLAKADADL 133
      75  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0b      66  EVRPVGSIVLNRNNTSGSDVQAGSLVQIDPATYQASVDSKGLAKESAAAIAHLTV 125
      67  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y      134  ARKPELVAAEAVSRQEYDAAVTAKESAEGVYKAAQAIKSGAININRSRTIAPISGFIOQ 193
      135  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0b      126  KRYPVLVGTXYISQGEYDQAIADANQADAAVAIAKATYESARINLAATYKTPISGRIGK 185
      127  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y      194  SKVSEGTLLNMGDTTVLATIROTNPMYNNVTOSASEVMKLRQIAEGKLLAADGVITANGI 253
      195  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0b      186  STVTEGALVNTGOETTELATVQOJLDEIYVDVTQSSNDFMRKQSVQGINLKHENATSNVEL 245
      187  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y      254  KFDGCTVYPEKGRLLFADPVVNVESTGOITLRAAVNDQNIIMPGLIYRVLMQVAVDAAF 313
      255  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0b      246  VMENQOTPELKGTLQFSDVTVDESTGSLTLRAVFPNPOHTLLPGMFPEARIDEGVQPAI 305
      247  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y      314  VVPQQAIVTRGAKD--TWMTLVNAGGEMERREVVVAQOQGINMTLVISGLKDGDGVNVEGSI 371
      315  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0b      306  LIPQGVSRTPRGDATVLIIVNKSQVEARPVVASQAIQGMKMLISEGLKSGDQIVISGL-- 363
      307  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y      372  AGITGAKVTVPKEMASSENQAAPOSGVQGTASEAKTASE 410
      373  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0b      364  -----QKARPE-----QVKATTDPADTAK 385

```

RESULT 16		
Q9F241		
ID	Q9F241	PRELIMINARY; PRT; 394 AA.
AC	Q9F241;	
DT	01-MAR-2001 (TrEMBLrel. 16, Created)	
DT	01-MAR-2001 (TrEMBLrel. 16, last sequence update)	
DT	01-OCT-2002 (TrEMBLrel. 22, last annotation update)	
DE	Putative membrane fusion protein.	
GN	SMD.	
OS	Xanthomonas maltophilia (Pseudomonas maltophilia) (Stenotrophomonas maltophilia).	
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;	
OC	Xanthomonadaceae; Stenotrophomonas.	
OX	NCBI_TaxID=40324;	
RM	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=D457R;	
RX	MEDLINE=20493115; Pubmed=11036026;	
RA	Alonso A., Martinez J.L.;	
RT	"Cloning and characterization of SmeDEF, a novel multidrug efflux pump	
RL	from <i>Stenotrophomonas maltophilia</i> ."	
	Antimicrob. Agents Chemother. 44:3079-3086(2000).	

DR EMBL; AJ252200; CAC14594.1; -.
DR InterPro; IPR006143; HLYD.
DR Pfam; PF005029; HLYD; 1.
DR SEQUENCE 394 AA; 40912 MW; D4234A8037E47B53 CRC64;

Query Match	36.8%;	Score 743.5;	DB 2;	Length 394;
Best Local Similarity	41.8%;	Pred. No. 2.6e-33;		
Matches 167;	Conservative 76;	Mismatches 144;	Indels 13;	Gaps 4;

```

QY      9 MRPAALAAAVLVLSGCKGKGDMAOGOPAGRREAPVAVGVTHPOTVALTVELPGRLE 68
Db      6 IRFPALSLIAATVAAC-----GGQPQAPBEGQPGGVTVTLKSTVGLTRRLPERTN 57
QY      69 SLETDVVRQVGGIIOKRLPQEGSYVRQGPVYQIDSTYEANLESAPQALATQATLAK 128
Db      58 APLVAERQVGVGIYAKRLFTGCGVWKAGPEPLYQLDDASYPAQNNNRAQOLAREAEATANA 117
QY      129 ADADLARMPVLAAEASRQEYDAVLTAKTSSEAGVNRQAOKALISAGINLNRSPITAPIS 188
Db      118 ARLSAKRTIELLAKVDVDSQODLENNVAQOKQLEADVGAAKSLDPAANVTLLGYATITAPIS 177
QY      189 GFLGQSKVSEGLTLNAGDITTVLATIRQTNPMVYVNTQSASEVMKLRLQIAEGKLLAADGV 248
Db      178 GRGKSSVTVGALVSAQGNALATATQQLDPYVDLTQSSABLLQRLRELAAGR-IDQNOT 226
QY      249 IANGIRFEDGTYYPEKGLLFPADPVYNASTQGITLRAAPVPDQNIILMPGLYVRLMDQVA 308
Db      237 LPLSILMEGGSFFEHNGTILFESEVSDPTTSGFLRYKRVNDPDLMLPMGYVAIVLGGGV 296
QY      309 VDAVAVVPOQAATVRGAK--DTVMIVNAGMGEPRELVTAQOQGTNMIVTSGLKDGDKRVVV 366
Db      297 RSPAVLVPMQGIARDPKQDPTTAMVVGKXNKYEVAPVYKSRVGBQXMLVEDGLKAGDKVIV 356
QY      367 EGISINGITGAKKVTPEKMASSENOAAAPOSGCVQTASEAK 406
Db      357 EGLQKIGPQMVPKATEKGDAPEKPAALAAQPPA--PAGAK 394

```

RESULT 17	PROTEIN	PRELIMINARY	PRT	404 AA.
ID	OEP101			
AC	OEP101			
DT	01-OCT-2002 (TrEMBLrel. 22, Created)			
DT	01-OCT-2002 (TrEMBLrel. 22, Last sequence update)			
DT	01-OCT-2002 (TrEMBLrel. 22, Last annotation update)			
DE	Multidrug resistance protein.			
GN	MEXA OR XAC2844.			
OS	Xanthomonas axonopodis (pv. citri).			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;			
OC	Xanthomonadaceae; Xanthomonas.			
OX	NCBI_TaxID=92829;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=306 / ATCC 13902 / XY 101;			
RX	MEDLINE=22022145; PubMed=12024217;			
RA	da Silva A.C.R., Ferro U.A., Reinach F.C., Farah C.S., Furjan L.R.,			
RA	Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,			
RA	Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,			
RA	Caracote G., Cannavaro F., Cardoso J., Chambergro F., Ciapina L.P.,			
RA	Ciavarella R.M.B., Coutinho L.L., Curisno-Santos J.R., El-Dorri H.,			
RA	Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,			
RA	Formighieri E.F., Franco M.C., Greggio C.C., Guber A.,			
RA	Katsunuma A.C., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,			
RA	Locali E.C., Machado M.A., Madalra A.M.B.N., Martinez-Rossi N.M.,			
RA	Martins E.C., Melandri J., Menck C.F.M., Miyaki C.Y., Moon D.H.,			
RA	Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,			
RA	Pereira H.A., Rossi A., Sena U.A.D., Silva C., de Souza R.F.,			
RA	Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,			
RA	Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,			
RA	Setubal J.C., Kitejima J.P.,			
RT	"Comparison of the genomes of two Xanthomonas pathogens with differing			
RT	host specificities."			
RL	Nature 417:459-463(2002).			

DR EMBL: AE011925; AAM37689.1; --
 DR InterPro: IPR006143; H1YD.
 DR Pfam: PF00529; H1YD.
 DR Complete proteome.
 KW SEQUENCE 404 AA; 42179 MW; F82FD7C04BC7C6B7 CRC64;

Query Match 36.7%; Score 741.5; DB 16; Length 404;
 Best Local Similarity 42.3%; Pred. No. 3.5e-33;
 Matches 174; Conservative 65; Mismatches 151; Indels 21; Gaps 7;

QY 8 AMRAAALAAVA--LVLSGCK--GGDAAGGQPRGRREAPVGVVTHPQVATLVEL 63
 DB 6 SFRSLVAVAATLALACSPGPGPPQEG-----TPKGVLTIVKQOPVLTTEL 57
 QY 64 PGRLESLTADVAOVGGIIQKRLFOEGSYVRAQOLYQIDSTYEANLESARQATATQ 123
 DB 58 PGRTPVYLSERPGVIGVSGROFTBGGDVKAGGLVQIDPAQVRASTASQASLAKQ 117
 QY 124 ATLAKADADLARKPLVAEAVSRQDYDAVTAKRSAEAGVKAQAIAKSAGININRSRI 183
 DB 118 ATLRTAQLKARRYKELAIKAISSQEGDDTDALQAKADVAAGKASVETARINLAFRL 177
 QY 184 TAPISGFGOSKVSSEGLTNGDTVLATIRQTNPMYVNVQASBWKRLQALBGL- 242
 DB 178 DAPISGRIGRSSVTFGALVTANQATALTIIQDLDTYIDVTQPSAVLRISQAMARGDL 237
 QY 243 LAADGVIAVGIKFDGTVYPERGRLFPADPVNVESTQITLRAAVPNDONTIMPLGLYRV 302
 DB 238 QAGGAARVSLLEBGSVYPIQGHAFSDVVDQNTGSIITLRAVFPNNABLLPMTYRA 297
 QY 303 LMDQVAVDAVFPQAAVTR--GAKDTVMIVNAQGMEPREVTAAQOGTNIWVTSGLKD 360
 DB 298 VLOEGVKEQGVILVPOAAVSRNAGKPTAFVVGADHKLQLRLAETREVDQMLVRSGLK 357
 QY 361 GDKVYVREGISITAGITGAKVTPKKEASSENQAAAPQSGVQTAASEKTTSEA 411
 DB 358 GEQVVEGASRARDGIVKTVF--W---QPKATPAAAGNAGPSAPATPRA 402

RESULT 18

Q8E8H1 PRELIMINARY; PRT; 382 AA.
 AC Q8E8H1;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Multidrug resistance protein, AcrA/AcrE family.
 GN S04693.
 OS Shewanella oneidensis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
 OC Alteromonadaceae; Shewanella.
 NX NCBI_TaxID=70863;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MR-1;
 RX MEDLINE=22297686; PubMed=12368813;
 RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
 Read T.D., Eisen J.A., Seehadri R., Ward N., Meehe B., Clayton R.A.,
 Meyer T., Tsapin A., Scott J., Beanan M., Brinkac J., Daugherty S.,
 Deboy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
 Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
 Vamathevan J., Weidman J., Imprim M., Lee K., Berry K., Lee C.,
 Mueller J., Khouri H., Gill J., Utlcherback T.R., McDonald L.A.,
 RA Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.;
 RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
 RT Shewanella oneidensis.";
 RL Nat. Biotechnol. 20:1118-1123(2002).
 DR EMBL: AE015901; AAM57652.1; --
 DR TIGR: S04693; --
 KW Complete proteome.
 SQ SEQUENCE 382 AA; 40186 MW; 2BBB85FC4DF09A53 CRC64;

Query Match 36.4%; Score 734; DB 16; Length 382;

Best Local Similarity 44.0%; Pred. No. 8.3e-33;
 Matches 168; Conservative 74; Mismatches 130; Indels 10; Gaps 6;

QY 13 ALAAVALVSSCGKGD-AAQGGQPRGRREAPVGVVTHPQVATLVELPGLSEIR 71
 DB 8 ASVISVALMMMAKSPQEKKAHGG--AGPQSE--VGVIKVEAKQVIKIELPGRSKAFL 63
 QY 72 TADVAQVGGIIQKRLFOEGSYVRAQOLYQIDSTYEANLESARQATATQATKADA 131
 DB 64 EAEVRPQVNGIITRSFPEGGNVKGESLYQIDATYKALVSNADLAKANASLASAKA 123
 QY 132 DLARKPLVAEAVSRQDYDAVTAKRSAEAGVKAQAIAKSAGININRSRIAPISGFI 191
 DB 124 KAARYQVLKTNALSKOPDEADAAVKALASVTAEBAAINTAKINILEYTEVLAPISGRI 183
 QY 192 GQSVSESTLNAAGDTVLATIRQTNPMYVNVQASBWKRLQALBGLAADGVATV 251
 DB 184 GKSSVTAGALVTANQATALTIIQDLDTYIDVTQPSAVLRISQAMARGDL-DV 242
 QY 252 GIKFDDGTVYPERGRLFPADPVNVESTQITLRAAVPNDONTIMPLGLYRVLMQVAVDN 311
 DB 243 QLVIVDGVYHTKGLQFAEVSVDQNTGSIITLRAVFPNNABLLPMTYRAVLAAGTDPQ 302
 QY 312 AFVVPQAAVTRGARD--TWMIVNAQGMEPREVTAAQOGTNIWVTSGLKDGKRVVGEI 369
 DB 303 AILVPKALITRNAGKEAVAMVNAQGVKAEKIVTTAEVYNQMRVTSGLAVGDVIVEGJ, 362
 QY 370 SIAGITGAKVTPKKEASSENQ 391
 DB 363 Q-KIRPGAPVIPVISAATQ 382

RESULT 19

Q9MWZ9 PRELIMINARY; . PRT; 384 AA.
 AC Q9MWZ9;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE TEGA.
 GN TTGA.
 OS Pseudomonas putida.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 NX NCBI_TaxID=303;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DOT-T1;
 RX MEDLINE=98317269; PubMed=9642183;
 RA Ramos J.L., Duque E., Godoy P., Segura A.;
 RT "Efflux pumps involved in toluene tolerance in Pseudomonas putida DOT-
 RT T1E.";
 RL J. Bacteriol. 180:3323-3329(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DOT-T1;
 RA Segura A., Duque E., Ramos J.L.;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF031417; AAD39553.1; --
 DR InterPro: IPR006143; H1YD.
 DR Pfam: PF00529; H1YD; 1.
 SQ SEQUENCE 384 AA; 41249 MW; 1E1P9BA412D0F6C4 CRC64;

Query Match 36.2%; Score 730.5; DB 2; Length 384;
 Best Local Similarity 40.2%; Pred. No. 1.3e-32;
 Matches 169; Conservative 80; Mismatches 120; Indels 51; Gaps 10;

QY 6 FK-AMRAAALAAVALVSSCGKGDAAQGGQPRGRREAPVGVVTHPQVATLVELP 64
 DB 3 FKRAVTRVAAVALATLISGCK-----EBAAPAAQAPQGVVTHPQVATLVELP 54
 QY 65 GRLESRTADVAOVGGIIQKRLFOEGSYVRAQOLYQIDSTYEANLESARQATATQ 124

Db 55 GRTSAVRVAEVRPQVNGIILKRLFKESGEVKGQQLYQIDPAVYEATLANKANL--A 111
 QY 125 TLAKADADLARYKPLVAEAASVROEYDAVTRAKSAEAGVKAAQAAIKSAGINLRSRIT 184
 Db 112 TRSLAE----RYKQIDEGQAVSKQRYDDA-NAKR-----LQAEKSLSAQIDLRFTYTVL 160
 QY 185 APISGFIFQSKVSEGTLLNAGDTVLATIRQTNPMYVNTQSASEVMKLRLQIAEKL-L 243
 Db 161 APISGRIRGRSSFTFEGALVSNQGTDMATIQQLDPYVDVSTQSTABLLKRLRDLSEGLQK 220
 QY 244 AADGVIAGIKFDDGTVPYKPKGRLLPADPVNNESTQITLRAAVNDQNIIMPGLYRVL 303
 Db 221 AGNNASVQVLVEDSGTLFKQEGRELFSEVADETTGVSVTLRALFPNPHTLLPGMFVHAR 280
 QY 304 MDQVAVDPAFVFPQQAATRGAK--DTWMIVNAQGMPEPREVTVAAQOQGTNMIIVTSGLDKG 361
 Db 281 LKAGVNAVAIILAPQGVTRDLKGAFTALVNVQENKVELRLKASTLGSMDLIEGLNPG 340
 QY 362 DKVVEG-----ISIAGITGAKKVTPEKMASSENQAAAPOSQVGTASEAKTASAE 412
 Db 341 DRLITGGLQYVRPGVEVTKVSDATNVK-----PAGPDQANAAKADAKAE 384

RESULT 20

Q8PQJ6 PRELIMINARY; PRT; 408 AA.
 AC Q8PQJ6; 08PQJ6;
 DT 01-OCT-2002 (TREMBlrel. 22, Created)
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Acridinyl resistance protein.
 GN SMDR OR XAC0327.
 OS Xanthomonas axonopodis (pv. citri).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xanthomonas.
 CX NCBI_TaxID=92829;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=306 / ATCC 13902 / XV 101;
 RX MEDLINE=22022145; PubMed=12024217;
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furian L.R.,
 Quagga R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
 Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.B.A.,
 Camaroto G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
 Cicarelli R.M.B., Coutinho L.V., Curcio-Santos J.R., El-Dorriy H.,
 Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
 Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
 Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
 Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
 Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
 Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
 R. Spindola L.A.F., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
 Trindade dos Santos M., Truffi D., Tsai S.M., White P.F.,
 RA Secubal J.C., Kitajima D.P.;
 RT "Comparison of the genomes of two Xanthomonas pathogens with differing
 host specificities";
 RL Nature 417:459-463(2002).
 DR EMBL; AB011658; AAC55219.1; -.
 DR InterPro; IPR006143; HLYD.
 DR Pfam; PF00529; HLYD; 1.
 KW Complete proteome.
 SQ SEQUENCE 408 AA; 42533 MW; 55DD3E271D691265 CRC64;

Query Match 36.0%; Score 727.5; DB 16; Length 408;
 Best Local Similarity 41.2%; Pred. No. 2.1e-32;
 Matches 165; Conservative 80; Mismatches 140; Indels 15; Gaps 6;

QY 14 LAAVAVLSSCGKGDAAAGGPGAGRAPAVGVVTVHQTVALTVLTLGRLESLRTA 73
 Db 18 LPAATVTLMAAC-----SSRQPP--QMPQTVGVVTLKQRLAVIDDTLGRVYAVYTS 68
 QY 74 DVRAQVGGIIQKRLFOEGSVYRAGQPLVQIDNSTVEANLESARAQLATAQTLAKADL 133

Db 69 DVRPQVGIVAKRLFTFGQDVQAOQVLYEIDPASYQAAVDTAKGDLAQAEAAVISAAPKA 128
 QY 134 ARYPLVAABASQCEYDAVTRAKSAEAGVKAAQAAIKSAGINLRSRITAPISGFIQ 193
 Db 129 ORYOTLVGLDVAISKQDDDLATIRSNBAVNAKASLQTRIMLDYTRITAPISGRIGT 188
 QY 194 SKVSEGTLLNAGDTVLATIRQTNPMYVNTQSASEVMKLRLQIAEKLAAADGVIAGI 253
 Db 189 SSTYSGALVSAQGEVLATINQDLPIYVDVTSQAQLQRLQDQACQLKAVDGAQVTL 248
 QY 254 KFDGTYTPKPKGRLLPADPVNNESTQITLRAAVNDQNIIMPGLYRVLMDQVAVDPAF 313
 Db 249 QLEDGSTYAHSGTLEVDPAADVATGTVKRAAVPNPNERLLPQMYVTAKLSMAVDEQAI 308
 QY 314 VVPOQAATRGAKD--TWMIVNAQGMPEPREVTVAAQOQGTNMIIVTSGLDKGKPVVEGISI 371
 Db 309 LVFPQQAISRSKGAVALVSGDNKVAQRVLITGDAIGDKVVRQGLKAGDKVIVQGLQK 368
 QY 372 AGITGA--KRVTPREMASSENQAA-APQSGVGTASEAKTA 408
 Db 369 ASV-GAEVKAVEVTPAAMAGNAATAPATQGTAAAPKA 407

RESULT 21

Q8ZC88 PRELIMINARY; PRT; 395 AA.
 AC Q8ZC88; 08ZC88;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Multidrug efflux protein (Acridine efflux pump).
 GN ACRA OR YPO3132 OR Y1050.
 OS Yersinia pestis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Yersinia.
 CX NCBI_TaxID=632;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CO-92 / Biovar Orientalis;
 RX MEDLINE=21470413; PubMed=11586360;
 RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
 Pritchett M.B., Sebahthia M., James K.D., Churcher C., Mungall K.L.,
 Baker S., Baaham D., Bentley S.D., Brooks K., Cerdano-Tarrega A.M.,
 Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
 Feltwell T., Hamlin N., Holroyd S., Jaseles K., Karlshew A.V.,
 Leather S., Moutie S., Oyston P.C.F., Quail M., Rutherford K.,
 Simmonds M., Skelton J., Stevens K., Whitehead S., Barrett B.G.,
 RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
 RL Nature 413:523-527(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KIM5 / Biovar Medinevalis;
 RX MEDLINE=22137863; PubMed=12142430;
 RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
 RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
 RA Herberston J.D., Lindler L.B., Brubaker R.R., Plano G.V.,
 RA Straley S.C., McDonough K.A., Nilles M.L., Watson J.S., Blattner F.R.,
 RA Petry R.D.;
 RT "Genome sequence of Yersinia pestis KIM.";
 RL J. Bacteriol. 184:4601-4611(2002).
 DR EMBL; AJ141415; CAC92367.1; -.
 DR EMBL; AB013708; AAC84631.1; -.
 DR InterPro; IPR006143; HLYD.
 DR Pfam; PF00529; HLYD; 1.
 KW Complete proteome.
 SQ SEQUENCE 395 AA; 41889 MW; 8D25E40E22ECBA36 CRC64;

Query Match 36.0%; Score 726.5; DB 16; Length 395;
 Best Local Similarity 40.6%; Pred. No. 2.2e-32;
 Matches 164; Conservative 73; Mismatches 142; Indels 23; Gaps 5;

QY 11 AALAAVAVLSSCGKGDAAAGGPGAGRAPAVGVVTVHQTVALTVLTLGRLESL 70

```

Db 11 AAILVSSGLVIGC-NDKDAQAQ-----AQAQAEVGVVILKAPLITITDLPRTISAF 64
Qy 71 RTADVRAVGGIITQRLFOEGSVYRAGOPLYOIDSSTEYANLESARQAATATQAATLAVAD 130
Db 65 RVAEVRPQVSGIILKRNATIEGSDVYRAGSLYVIDPATTOAAYDSKAGDLAKQAQSAQIAH 124
Qy 131 ADLARYKPLVAEAVEROEYDAAVTAKRSAAEYVKAQAQAIAKSAGININRSRTAPISGF 190
Db 125 LTVNRYKPLLGTYNYSKOEYDQALSDAQADATVLAAKAALLESANINLAAYQVRSPIGR 184
Qy 191 IGOSKVSSEGTLLNAGDTTVALITRQTNPMYVNTQSASEWKLKROIAEKKLLAADGVIA 250
Db 185 TGKSAVTGALVTSQAASAMTVOQLDPMYVDVTSQSDERFLKKEKELADGILKQENGRKX 244
Qy 251 VGIKFDDGTIVPEKGRLLFADPVNVESTGQITLRAAVPNQDNIIMPGLYVRVLMDOQAVD 310
Db 245 VRLLENGVETETITLSPSGVTVDITGSLTRALFPNPEALDPGFVRDLDEGIRP 304
Qy 311 NAFVVPQQAQVTRGAKD--TVMIVNAQGMPEPRTVTAQOOGTMTVTSGLKDGDKVVEG 368
Db 305 DALVFPQQAQVTRNPRGEGATAVVGAQNKVEMKRTLLANQAIGKMLVTGGLKAGDRLLISG 364
Qy 369 ISIAGITAKKVTTPKEMASSENQAAPQSGVOTPAASEAKTASEAE 412
Db 365 L-----OKIKP-----GVEVKVQEVDTIDTAPETAPADTAK 393

```

RESULT 22

Q8UB14 PRELIMINARY; PRT; 427 AA.

```

AC Q8UB14;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE RND multidrug efflux membrane permease.
GN ACRA OR ATU3203 OR AGR_L_3215.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eissen J.A., Karp P.D., Boye D. St.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58."
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Ounjillo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmiel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Woliam C., Allinger M., Doughy D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursen J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RL Science 294:2323-2328(2001).
DR EMBL; AF009250; AAL44019.1; ALT_INIT.
DR EMBL; AB008363; AAK90184.1; -.
DR InterPro; IPR006143; HLYD.
DR Pfam; PF00529; HLYD; 1.

```

KW Complete proteome.
SQ SEQUENCE 427 AA; 45728 MW; 9E7C2E07AC8B9781 CRC64;

Query Match 35.8%; Score 723; DB 16; Length 427;
Best Local Similarity 40.0%; Pred. No. 3.9e-32;
Matches 161; Conservative 83; Mismatches 143; Indels 16; Gaps 7;

```

Qy 11 AALAAVALVSSCGKGDAAQGGOPAGREAPAP-VGVVTVHPQTVALTVELPGRLES 69
Db 32 AAVLTGIVLV-----GCSDEQASAP---APPGAVKVAVAVPELPIINELPGRAP 82
Qy 70 LRTADVRAQVGGIITQRLFOEGSVYRAGOPLYOIDSSTEYANLESARQAATATQAATLAKA 129
Db 83 TRLEAVPRVSGIILKRNATIEGSDVYRAGSLYVIDPATTOAAYDSKAGDLAKQAQSAQIAH 142
Qy 130 DADLARYKPLVAEAVEROEYDAAVTAKRSAAEYVKAQAQAIAKSAGININRSRTAPISG 189
Db 143 RQTDARQOQLRRSNVSGOQEPDNALIALQAQDAEVAVAEAGVAERMLQYADVAPAPISG 202
Qy 190 FIGOSKVSSEGTLLNAGDTTVALITRQTNPMYVNTQSASEWKLKROIAEKKLLAADGVI 249
Db 203 VIGARITGEGALVATGSENATITQQLDPIYADTQPRADIRLRKALQDQQLMTGQNEA 262
Qy 250 AVGIKFDGTIVPEKGRLLFADPVNVESTGQITLRAAVPNQDNIIMPGLYVRVLMDOQAV 309
Db 263 EVNLIFPDGSRYPVSGRLFSEAADVDTTGQVTLRGEPFNPNGDLLPMYVYRVOIQGSIQ 322
Qy 310 DNAFVVPQQAQVTR--GAKDTVMIVNAQGMPEPRTVTAQOOGTMTVTSGLKDGDKVVE 367
Db 323 KAFAVFPQQAQVTRGAGGASVLYVNAEDTVEQRRSVSGRSIGDVISEGDDDDRVAAE 382
Qy 368 GISIAGITAKKVTTPKEMASSENQAAPQSGVOTPAASEAKTASE 410
Db 383 GQKTA-PGA-KVPEPMSQEPDVAAL--AGSEGAASEISSE 421

```

RESULT 23

Q9KJC3 PRELIMINARY; PRT; 371 AA.

```

AC Q9KJC3;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Periplasmic linker protein.
GN APPA
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=S12;
RA MEDLINE=21097242; PubMed=11160799;
RA Kleboom J., de Bont J.A.M.;
RT "Identification and molecular characterization of an efflux system
RT involved in Pseudomonas putida S12 multidrug resistance."
RL Microbiology 147:43-51(2001).
DR EMBL; AF183959; AAF73831.1; -.
DR InterPro; IPR006143; HLYD.
DR Pfam; PF00529; HLYD; 1.
SQ SEQUENCE 371 AA; 40276 MW; B86AACB9A6133645 CRC64;

```

Query Match 35.7%; Score 720; DB 2; Length 371;
Best Local Similarity 43.5%; Pred. No. 4.7e-32;
Matches 160; Conservative 74; Mismatches 108; Indels 26; Gaps 8;

```

Qy 6 FK-AMRAAALAAVALVSSCGKGDAAQGGOPAGREAPAPVGVVTVHPQTVALTVELP 64
Db 3 FKAVATLVSAVALATLISGCK-----EBAAPAAQAPQGVVTVITQQAFTLSLPL 54
Qy 65 GRLESRTADVRAQVGGIITQRLFOEGSVYRAGOPLYOIDSSTEYANLESARQAATATQA 124
Db 55 GRTSAVYVAEVRPQVNGIILKRLFKEGSEVKEGQQLYQIDPAVEATLAAKAKALL--A 111

```

Qy	125	TLAKADALARKPPLVAAEVARSROEYDAVTKKRSMAEYKAAQAARKSAGINLNSRTT	18
		: : : : : : : : : : : : : : : : : : : : : : : : :	
Db	112	TRSLAE-----RYKQULDEOAVSKOEYDDA-NKR-----LQEHASKRQIDRLRTKYL	160
Qy	185	APISGFIQSQKVSSEGTLLNAGDTVLTATIRQTNPMYVNTQSASEVWKLRQIOAEKUL-L	243
		: : : : : : : : : : : : : : : : : : : : : : : : :	
Db	161	APISGRIGRSSPTEGALVNSCGTDMATIQQLDPIVVDVTQSTAEILKLRRLDIESQLOK	220
Qy	244	AADGVIAIGIKRPPDCTVYBEKRLLPADPVNVESTQITLRAAPVPMQDMLMPGLVVRVL	303
		: : : : : : : : : : : : : : : : : : : : : : : : :	
Db	221	AGNNAASQVLVEDSLFKQBSRLSESEVAVBETTSVTLRALFPDPDHTLPLGMFVHAR	280
Qy	304	MDOVAVDAFVFPQQAATVRGAK--DTVMVNAQGGNEPREVTVAQOQGTNMTVTSGLKDG	361
		: : : : : : : : : : : : : : : : : : : : : : : : :	
Db	281	LKAGVANAHAIIAPQCGVTRDELKGAPFALVVGNEKNKVELRQLKASRTLSGDMILIEGILNG	340
Qy	362	DKVVEGI 369	
		: : : : :	
Db	341	DRLITEGL 348	

RESULT 24	
Q92T02	
ID Q92T02	PRELIMINARY; PRT; 407 AA

DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, last annotation update)
 DE Putative multidrug efflux system protein.
 GN R00193 OR SMC02868.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Sinorhizobium.
 NCBI_TaxId=382;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=21396507; Pubmed=11481430;
 RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
 RA Boistard P., Becker A., Boutmy M., Cadieu E., Dreaano S., Gloux S.,
 RA Gogrie T., Goffeau A., Kahn D., Kiss E., Lelaurie V., Masny D.,
 RA Pohl T., Portetalle D., Puchler A., Purnelle B., Ramsperger U.,
 RA Renard C., Thebaud P., Vandebol M., Weidner S., Gallbert F.,
 RT "Analysis of the chromosome sequence of the legume symbiont
 RT Sinorhizobium meliloti strain 1021."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882 (2001).
 DR EMBL; AL591782; CAC1580.1; -.
 DR InterPro: IPR006143; HlyD.
 KW pfam; PF00552; HlyD. 1.
 KW Complete proteome.
 SQ SEQUENCE 407 AA; 43732 MW; 53FAB3B0FC55042A CRC64;

Query Match	35.6%;	Score 719.5;	DB 16;	Length 407;
Best Local Similarity	40.2%;	Pred. No. 5.7e-32;		
Matches 164;	Conservative 79;	Mismatches 148;	Indels 17;	Gaps 71;

```
QY 9 MKAALAAVAVY - -LSGCGGDDAAQGGPAREKAPAP -VGVVTVHPOTVALTYELPG 65
Db 7 MNPRLIAAALVATYIFLAGCQCKNE - - - - -QAAAAAPPSPVAVFTTKAEPRLPTNELPG 61
QY 66 RLESRTADVRAQVGIIQKRLFOEGSSVYRAGOPLOQDSSTEANLESRAOLAATQAOT 125
Db 62 RITAIRLAEVRRISIGIIVERVEFGQTMWEGSVLVRIDPAPQVKYDSAEATLKRAQAV 121
QY 126 LAKADADLARVYKPLVAAEAVSRQEDYDAVTAKESAENGVAKAQDAIKSAGININRSRITA 185
Db 122 VDAQARTADROGRSLKEAQVTAQVQYDDALAAALQAADVGIABAGLAELKNTQYTWTA 181
QY 166 PISGFGGSGKVBEGTLLNAGDTTVLATIKOTNPMYVNTQSAEVMKLRQIABGKLLAA 245
Db 182 PISGRGRALLREGALVNTNDPONLANTIQOLDIYADFQSSADLRLRLRAKLDGQWMSA 241
```

Qy	246	DGVIANGIKEDDQYTPYEGKRLLEPVANSTGCIITRAAPVNDQIILNPLGLVRLVMD	305
		: : : : : : : : : : : : : : : : : : : : : : : : :	
Db	242	KNEAEVQLLLDGTGPALKGRLLFSEAAVDATTQVTLRGEFPFPNNDDLPGHIVRVOIQ	301
		: : : : : : : : : : : : : : : : : : : : : : : : :	
Qy	306	QYAVNDAPFVPOOAATVR--GAKDTWMIYNAQCGMEPREVTYAAQCGTNMTIVTSGLKXGDK	363
		: : : : : : : : : : : : : : : : : : : : : : : : :	
Db	302	QGLEKAITVPOAAVORNNAGOSQVYVNAANNKYEFRRVTLIGRVGGRMÖVTSGLKXGEEK	361
		: : : : : : : : : : : : : : : : : : : : : : : : :	
Qy	364	VVVEGISIAGITGAKKVTPEK-----ASSENQAAAPDSQVOTASEAK	406
		: : : : : : : : : : : : : : : : : : : : : : : : :	
Db	362	VIVYEGQKVG-PGA-PVQPSMDPAAKAPAPQASASADAGKAPATYK	407

RESULT 25	
Q8GC84	
ID Q8GC84	PRELIMINARY; PRT; 374 AA

DT	01-MAR-2003	(TREMblrel. 23, Created)	
DT	01-MAR-2003	(TREMblrel. 23, Last sequence update)	
DT	01-MAR-2003	(TREMblrel. 23, Last annotation update)	
DE	DefA lipoprotein precursor.		
GN	BEFA.		
OS	Enterobacter aerogenes (Aerobacter aerogenes).		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
OC	Enterobacteriaceae; Enterobacter.		
OX	NCBI_TaxID=548;		
RN	[1]		
RX	SEQUENCE FROM N.A.		
RP	STRAIN=BM16627;		
RC			
RA	Masl M., Pages J.M., Pradel E.,		
RT	"Identification and characterization of the Enterobacter aerogenes		
RT	eefAa operon encoding a putative tripartite efflux system.";		
RL	Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: AJ508047; CAD8861.1; -.		
FT	Signal; Lipoprotein.		
FT	STAIN	1	17
FT	CHAIN	18	374
QO	SEQUENCE	374 AA; 39374 MW; 3328FD6AFB942BB CRC64;	POTENTIAL. BEFA LIPOPROTEIN.

Query Match 35.6%; Score 719; DB 2; Length 374;

Matches 164; Conservative 65; Mismatches 122; Indels 22; Gaps 7;

12 AATAAVALVSSCGKGGAQGQPA GREAPAPVGVTVHPQTVALTVELPGRLESLR 71.

6 APIATAALFL-LSGC-----DNVQTAP---QQTPEVGIVTLQAQPPVPVSQLTGRTTASL 57

72 TADVRAOVGGITOKRLEOEGSYVRAGOPLOYIDSSTYEANLESARAOLATAOATLAKADA 131

58 SAEVBPVGGITOKRLFTEGDMYKAGOALYOIDPSSYRATWNEAAALKQAOALVSDCO 117

132 PIAPYKPTVAAEAVSPQEVDAVVTAKRSAEAGVKAQQAATKSAGINLNRSBITAPISGET 191

119 KLOBYASIPDNQVSPDADNAA\$ACAN\$KASVESKKAAL\$ESABTNI\$NTWTTV\$TAPIAGBT 177

100 СОСТАВЛЯЕТ ИСТОЧНИКОВЫЙ МАТЕРИАЛ ПРОЕКТА ЗАДАЧА

[illegible]

0 1 2 3 4 5 6 7 8 9 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z [\] ^ _ ` a b c d e f g h i j k l m n o p q r s t u v w x y z { | } ~

2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040 2041 2042 2043 2044 2045 2046 2047 2048 2049 2050 2051 2052 2053 2054 2055 2056 2057 2058 2059 2060 2061 2062 2063 2064 2065 2066 2067 2068 2069 2070 2071 2072 2073 2074 2075 2076 2077 2078 2079 2080 2081 2082 2083 2084 2085 2086 2087 2088 2089 2090 2091 2092 2093 2094 2095 2096 2097 2098 2099 2100 2101 2102 2103 2104 2105 2106 2107 2108 2109 2110 2111 2112 2113 2114 2115 2116 2117 2118 2119 2120 2121 2122 2123 2124 2125 2126 2127 2128 2129 2130 2131 2132 2133 2134 2135 2136 2137 2138 2139 2140 2141 2142 2143 2144 2145 2146 2147 2148 2149 2150 2151 2152 2153 2154 2155 2156 2157 2158 2159 2160 2161 2162 2163 2164 2165 2166 2167 2168 2169 2170 2171 2172 2173 2174 2175 2176 2177 2178 2179 2180 2181 2182 2183 2184 2185 2186 2187 2188 2189 2190 2191 2192 2193 2194 2195 2196 2197 2198 2199 2200 2201 2202 2203 2204 2205 2206 2207 2208 2209 2210 2211 2212 2213 2214 2215 2216 2217 2218 2219 2220 2221 2222 2223 2224 2225 2226 2227 2228 2229 2230 2231 2232 2233 2234 2235 2236 2237 2238 2239 2240 2241 2242 2243 2244 2245 2246 2247 2248 2249 2250 2251 2252 2253 2254 2255 2256 2257 2258 2259 2260 2261 2262 2263 2264 2265 2266 2267 2268 2269 2270 2271 2272 2273 2274 2275 2276 2277 2278 2279 2280 2281 2282 2283 2284 2285 2286 2287 2288 2289 2290 2291 2292 2293 2294 2295 2296 2297 2298 2299 2300 2301 2302 2303 2304 2305 2306 2307 2308 2309 2310 2311 2312 2313 2314 2315 2316 2317 2318 2319 2320 2321 2322 2323 2324 2325 2326 2327 2328 2329 2330 2331 2332 2333 2334 2335 2336 2337 2338 2339 2340 2341 2342 2343 2344 2345 2346 2347 2348 2349 2350 2351 2352 2353 2354 2355 2356 2357 2358 2359 2360 2361 2362 2363 2364 2365 2366 2367 2368 2369 2370 2371 2372 2373 2374 2375 2376 2377 2378 2379 2380 2381 2382 2383 2384 2385 2386 2387 2388 2389 2390 2391 2392 2393 2394 2395 2396 2397 2398 2399 2400 2401 2402 2403 2404 2405 2406 2407 2408 2409 2410 2411 2412 2413 2414 2415 2416 2417 2418 2419 2420 2421 2422 2423 2424 2425 2426 2427 2428 2429 2430 2431 2432 2433 2434 2435 2436 2437 2438 2439 2440 2441 2442 2443 2444 2445 2446 2447 2448 2449 2450 2451 2452 2453 2454 2455 2456 2457 2458 2459 2460 2461 2462 2463 2464 2465 2466 2467 2468 2469 2470 2471 2472 2473 2474 2475 2476 2477 2478 2479 2480 2481 2482 2483 2484 2485 2486 2487 2488 2489 2490 2491 2492 2493 2494 2495 2496 2497 2498 2499 2500 2501 2502 2503 2504 2505 2506 2507 2508 2509 2510 2511 2512 2513 2514 2515 2516 2517 2518 2519 2520 2521 2522 2523 2524 2525 2526 2527 2528 2529 2530 2531 2532 2533 2534 2535 2536 2537 2538 2539 2540 2541 2542 2543 2544 2545 2546 2547 2548 2549 2550 2551 2552 2553 2554 2555 2556 2557 2558 2559 2560 2561 2562 2563 2564 2565 2566 2567 2568 2569 2570 2571 2572 2573 2574 2575 2576 2577 2578 2579 2580 2581 2582 2583 2584 2585 2586 2587 2588 2589 2590 2591 2592 2593 2594 2595 2596 2597 2598 2599 2600 2601 2602 2603 2604 2605 2606 2607 2608 2609 2610 2611 2612 2613 2614 2615 2616 2617 2618 2619 2620 2621 2622 2623 2624 2625 2626 2627 2628 2629 2630 2631 2632 2633 2634 2635 2636 2637 2638 2639 2640 2641 2642 2643 2644 2645 2646 2647 2648 2649 2650 2651 2652 2653 2654 2655 2656 2657 2658 2659 2660 2661 2662 2663 2664 2665 2666 2667 2668 2669 2670 2671 2672 2673 2674 2675 2676 2677 2678 2679 2680 2681 2682 2683 2684 2685 2686 2687 2688 2689 2690 2691 2692 2693 2694 2695 2696 2697 2698 2699 2700 2701 2702 2703 2704 2705 2706 2707 2708 2709 2710 2711 2712 2713 2714 2715 2716 2717 2718 2719 2720 2721 2722 2723 2724 2725 2726 2727 2728 2729 2730 2731 2732 2733 2734 2735 2736 2737 2738 2739 2740 2741 2742 2743 2744 2745 2746 2747 2748 2749 2750 2751 2752 2753 2754 2755 2756 2757 2758 2759 2760 2761 2762 2763 2764 2765 2766 2767 2768 2769 2770 2771 2772 2773 2774 2775 2776 2777 2778 2779 2780 2781 2782 2783 2784 2785 2786 2787 2788 2789 2790 2791 2792 2793 2794 2795 2796 2797 2798 2799 2800 2801 2802 2803 2804 2805 2806 2807 2808 2809 2810 2811 2812 2813 2814 2815 2816 2817 2

QUESTION

1.
 2.
 3.
 4.
 5.
 6.
 7.
 8.
 9.
 10.
 11.
 12.
 13.
 14.
 15.
 16.
 17.
 18.
 19.
 20.
 21.
 22.
 23.
 24.
 25.
 26.
 27.
 28.
 29.
 30.
 31.
 32.
 33.
 34.
 35.
 36.
 37.
 38.
 39.
 40.
 41.
 42.
 43.
 44.
 45.
 46.
 47.
 48.
 49.
 50.
 51.
 52.
 53.
 54.
 55.
 56.
 57.
 58.
 59.
 60.
 61.
 62.
 63.
 64.
 65.
 66.
 67.
 68.
 69.
 70.
 71.
 72.
 73.
 74.
 75.
 76.
 77.
 78.
 79.
 80.
 81.
 82.
 83.
 84.
 85.
 86.
 87.
 88.
 89.
 90.
 91.
 92.
 93.
 94.
 95.
 96.
 97.
 98.
 99.
 100.
 101.
 102.
 103.
 104.
 105.
 106.
 107.
 108.
 109.
 110.
 111.
 112.
 113.
 114.
 115.
 116.
 117.
 118.
 119.
 120.
 121.
 122.
 123.
 124.
 125.
 126.
 127.
 128.
 129.
 130.
 131.
 132.
 133.
 134.
 135.
 136.
 137.
 138.
 139.
 140.
 141.
 142.
 143.
 144.
 145.
 146.
 147.
 148.
 149.
 150.
 151.
 152.
 153.
 154.
 155.
 156.
 157.
 158.
 159.
 160.
 161.
 162.
 163.
 164.
 165.
 166.
 167.
 168.
 169.
 170.
 171.
 172.
 173.
 174.
 175.
 176.
 177.
 178.
 179.
 180.
 181.
 182.
 183.
 184.
 185.
 186.
 187.
 188.
 189.
 190.
 191.
 192.
 193.
 194.
 195.
 196.
 197.
 198.
 199.
 200.
 201.
 202.
 203.
 204.
 205.
 206.
 207.
 208.
 209.
 210.
 211.
 212.
 213.
 214.
 215.
 216.
 217.
 218.
 219.
 220.
 221.

[illegible]

2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526

RESULT 26

Q8CM42
ID Q8CM42 PRELIMINARY; PRT; 384 AA.
AC Q8CM42;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Acridavine resistance protein A precursor.
GN C1761.
OS *Escherichia coli* O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; *Escherichia*.
OX NCBI_TaxID=217992;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / ATCC 700928;
RX MEDLINE=23388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Raeko D., Buckles E.L., Lion S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.,
RT "Extensive mosaic structure revealed by the complete genome sequence
of uropathogenic *Escherichia coli*."
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AB016760; AANB0227.1; -.
KW Complete proteome.
SQ SEQUENCE 384 AA; 41129 MW; FC9E69E4583EC7B8 CRC64;
Query Match 34.7%; Score 701.5; DB 16; Length 384;
Best Local Similarity 42.5%; Pred. No. 5.1e-31;
Matches 162; Conservative 69; Mismatches 133; Indels 17; Gaps 7;
QY 7 KAMRAALAAVALVLSGCKGDAAGGQAPAGREAPAPVGVVTPVATLVELPGR 66
DB 10 KIMKYIAFSVYAMLLSGC-----DNTQSNNSPSETE---VGVVTKSQPVSVSELTR 62
QY 67 LESRTADVRAQVGGITQKLFQEGSYVRAGQPLVQIDSSTYEANLSARAQATL 126
DB 63 TSAALSLEVRPQVGGIIOKRLFKEGDLVKAQPLVQIDASQYAAWAEARALQAOA-L 121
QY 127 AKAAADLA-RKPLVAEAVSRQYDAVATKRSABGVKAAQAIKSAGINLRSRITA 185
DB 122 VKAOCQKQRTYRLVKENGVSQDADADQSTCAODKASVEKKALLETARINLDMVTYTA 181
QY 186 PISGFIQSKVSEGLTLNAGDTVLAITRTQNPVYVNTQASAEVMLKRIAGKLLAA 245
DB 182 PISGRIGISSTVPALVYASODTALTIRGLDMYVDITRSSVDLIRKQ-----SLATN 237
QY 246 DGVIAVGIKEDDGTVPYKGRLLPADPVNVESTQITLRAAVPNDONILMGLYRVLM 305
DB 238 SDTWSVSLIEDGTTYSKGRLELVEAVVDESTGVTLRATFPVPOQLLPGMVRAVD 297
QY 306 QVAVDANFVVPQAVTRGAD--TVMIVNAQGMPEPEVYVAAQGGTNMIVTSGKDGK 363
DB 298 EGWVEDAILRPOQGVTRDAKNANALVYVNDKNVEORTLETGETYGDKWLNLNGLHNGR 357
QY 364 VVEGISIAGITGAKKVTPE 384
DB 358 LIVEG--SAKVTSGQYKAVE 376

RESULT 27
Q8RSM2
ID Q8RSM2 PRELIMINARY; PRT; 388 AA.
AC Q8RSM2;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE MexC protease.
GN MEXC.
OS uncultured bacterium.
OC plasmid pB4.
OC Bacteria; environmental samples.
OX NCBI_TaxID=71133;

RN [1]
RP SEQUENCE FROM N.A.
RA Tauch A.;
RT "The Incp beta plasmid pB4 encodes a tripartite antibiotic efflux
system of the RND-MFP-OEP type conferring erythromycin and
roxithromycin resistance in *Pseudomonas* sp. B13."
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ431260; CAD24338.1; -.
DR InterPro; IPR006143; HLYD.
DR Pfam; PF00529; HLYD; 1.
KW plasmid.
SQ SEQUENCE 388 AA; 41104 MW; 83CE3D1E04BF737 CRC64;
Query Match 34.7%; Score 700.5; DB 2; Length 388;
Best Local Similarity 40.6%; Pred. No. 5.9e-31;
Matches 159; Conservative 81; Mismatches 111; Indels 21; Gaps 8;
QY 12 AALAAVALVLSGCKGDAAGGQAPAG--EAPAPVGVVTPVATLVELPGRLES 69
DB 11 SVISLVITMTLAGC-----DQPAEWGEAPPREVDVLAVKTEPFTVIALPGRIRP 61
QY 70 LRTADVRAQVGGITQKLFQEGSYVRAGQPLVQIDSSTYEANLSARAQATLAKA 129
DB 62 VRAVEARAVAGVLSRKFEBGADVKAQDVFLIDPAPFAAUSRAEQLAQADALFQT 121
QY 130 DADLARYKPLVAEAVSRQYDAVATKRSABGVKAAQAIKSAGINLRSRITAPISG 189
DB 122 QATVHRYEPLVKIDAVRQDPVATALQSKAKKRAQADLETARLDLGATVRAPIAG 181
QY 190 FIGQKVSSEGLTLNAGDTVLAITRTQNPVYVNTQASAEVMLKRIAGKLLADG-V 248
DB 182 RIGRAQTEGALVQGGSTLLARIQDLPVADFTQPAADALRLRAIAGKVGSDGKA 241
QY 249 IANGIKEDDGTVPYKGRLLPADPVNVESTQITLRAAVPNDONILMGLYRVLMQVA 308
DB 242 LLSLSV--DQTSIESRGLFTFTVYDRSTQIALRGRFNPQGVLLPGMVVRRTPQGN 298
QY 309 VDNAFVVPQAVTRG--AKDTVMIVNAQGMPEPEVYVAAQGGTNMIVTSGKDGKRVV 366
DB 299 SD-AIVLPQAVLRAIDGQASVWVLGAEINVEARPYTTGMQGARWQITBGLKGDVIV 357
QY 367 EGISIAGITGAKKVTPE-EWASSENQAAPOS 397
DB 358 S--SLAALQPGVTVVPRQGAQAPAAASQPA 387

RESULT 28
Q51395
ID Q51395 PRELIMINARY; PRT; 387 AA.
AC Q51395;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Membrane fusion protein MEXC (RND multidrug efflux operon fusion
protein MEXC).
GN MEXC OR PA4599.
OS *Pseudomonas aeruginosa*.
OC Bacteria; Proteobacteria; Gammaproteobacteria; *Pseudomonadales*.
OC Pseudomonadaceae; *Pseudomonas*.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PAO;
RX MEDLINE=97032139; PubMed=8878035; Zhao Q., Wada A., Yamasaki T.,
RA Poole K., Gotch N., Tsujimoto H., Li X.Z., Nishino T.,
RA Nishat S., Yamagishi J., Li X.Z., Nishino T.,
RT "Overexpression of the mexC-mexD-opr efflux operon in nfxB-type
multidrug-resistant strains of *Pseudomonas aeruginosa*."
RL Mol. Microbiol. 21:713-724(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAO1;
RX MEDLINE=20437337; PubMed=10984043;

RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody K.L., Coulter S.N., Folger K.R., Kas A., Lartig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Mong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reiter J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.,
 RT "Complete genome sequence of *Pseudomonas aeruginosa* PA01, an
 RT opportunistic pathogen.";
 RL Nature 406:959-964(2000).
 DR EMBL: U57969; ABA41956.1; -
 DR EMBL: AE004873; AAG07987.1; -
 DR InterPro: IPR006143; HLYD.
 DR Pfam: PF00529; HLYD; 1.
 DR Complete proteome.
 SQ SEQUENCE 387 AA; 40838 MW; 8CF75BF3B2314A0A CRC64;

Query/Match 34.6%; Score 699.5; DB 16; Length 387;
 Best Local Similarity 41.5%; Pred. No. 6,6e-31;
 Matches 162; Conservative 82; Mismatches 131; Indels 15; Gaps 9;

QY 10 RAALAAVAALVLSGCKGDAAGGAPAGREAPAPVGVTVHPQTVALTVLPGRLES 69
 Db 9 RIGALAAIA-LACG-PAERQEA-EMVLP-VEVLTVQAEPLLSSELPGRLEP 60
 QY 70 LRTADVRAOVGGIIQRLFOEGSYVRAGPOLYQIDSTYEANLESARAQLATQAQTLAKA 129
 Db 61 VRVAEVRARVAGIVVRKFEAGADYKAGDLFQIDPAPLKAASRABELANRAVLREA 120
 QY 130 DADLARIYPLVAEVSROEYDAVTAKRSAEAGKAAQAAIKSGINLNSRTAPIISG 189
 Db 121 QARVRYEPLVKIQAVSQODPDTATADRSAAARSQAQDLERLNLGVASVTAPISG 180
 QY 190 FIGOSKEGTLNAGDTTVLATTIRQTNPMYVNTQSAEWMKLRQIAEGLLAADQVI 249
 Db 181 RIGRLVTEGALVGGSEATLMARIQDLPITADFTQTAELRLDAIKGTLAAGDSQ- 239
 QY 250 AVGIKFDGTVYPEKGRLLFADPVVNESTGQITTLRAVPNDONILMPGLYVRVLMDOYAV 309
 Db 240 ALTLKRV-EGTYERGGALQFADVAVDRTGQIALKGFANPDGVLTPGMVYRVRRPQID 298
 QY 310 DNAPFVPOOAVTRGAKDT--VMIVAAQGMPEPREVTVAOQOGTNMTVTSGLKDGKVV 367
 Db 299 NQAILVPPRAVHRSSDGAQVWVAVADRAERSRVGTGWSRQITGEGLEPGRVIVG 358
 QY 368 GISISAGITGAKKVTPEK-EMASSENQAAAPQ 396
 Db 359 G--LAAVOPGVKIVKPKPGGAQAQAQSPAPQ 386

RESULT 29

O8X7E1

PRELIMINARY; PRT; 373 AA.

AC O8X7E1 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Putative efflux pump (Putative multidrug-efflux transport
 DE protein).
 GN Z2509 OR ECS1863.
 OS *Escherichia coli* O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller U.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;

RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.";
 RL Nature 409:529-533(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,
 RA Kubara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 DR EMBL: AE005377; AAG56521.1; -
 DR EMBL: AP002556; BAB35286.1; -
 DR InterPro: IPR006143; HLYD.
 DR Pfam: PF00529; HLYD; 1.
 DR Complete proteome.
 SQ SEQUENCE 373 AA; 39612 MW; 2A573537AB44CDBF CRC64;

Query/Match 34.5%; Score 697.5; DB 16; Length 373;
 Best Local Similarity 42.7%; Pred. No. 8.1e-31;
 Matches 162; Conservative 66; Mismatches 134; Indels 17; Gaps 7;

QY 9 MRAALAAVAALVLSGCKGDAAGGAPAGREAPAPVGVTVHPQTVALTVLPGRLE 68
 Db 1 MKYIATSVMLLLSG-DMTQSNSSPSERE--VGVVTKSGPVSVSLTRTS 53
 QY 69 SLRTADVRAOVGGIIQRLFOEGSYVRAGPOLYQIDSTYEANLESARAQLATQAQTLAK 128
 Db 54 AALSAEVRPOVGGIIQRLFKFEGDLVYAGPOLYQIDAAVQAAMNERRAALQQA-LVK 112
 QY 129 ADADLA-RYRPLVAEVSROEYDAVTAKRSAEAGKAAQAAIKSGINLNSRTAPI 187
 Db 113 ADCKAKRVARLVRKENVSGQDDADQSTCAQDQASVAAKKAALERLNLDMVTAPI 172
 QY 188 SGFIGOSKEGTLNAGDTTVLATTIRQTNPMYVNTQSAEWMKLRQIAEGLLAADG 247
 Db 173 SGRIGISVTPGALVTVASQDTALTITRGLDTPVVDLRRSSVDLLRLRKQ----SLATNSD 228
 QY 248 VIANGIFPDGTVYPEKGRLLFADPVVNESTGQITTLRAVPNDONILMPGLYVRVLMDOY 307
 Db 229 TMSVSLILDEGTVYSEKGRLELFEVAVDESGVTLEAIIPNPPOOQLPMPFARARDEG 288
 QY 308 AVDAFVPPQOAVTRGAKD--TWIVAAQGMPEPREVTVAOQOGTNMTVTSGLKDGKVV 365
 Db 289 VMDPAILAPQOQVTRDKAGNATLAVKXNKVQORTLETGETYGDKKVLVNGLSGDRLI 348
 QY 366 VEGISAGITGAKKVTPEK 384
 Db 349 VEG--SAKVTSGQTVKAVE 365

RESULT 30

O93K41

PRELIMINARY; PRT; 398 AA.

AC O93K41 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Acra protein.
 GN ACRA.
 OS *Klebsiella pneumoniae*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Klebsiella.
 OX NCBI_TaxID=573;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K23;
 RA Domenech-Sanchez A., Alberti S., Martinez-Martinez L., Garcia I.,
 RA Benedi V.J.;
 RT "Characterization and role in antimicrobial resistance of the AcraB
 RT efflux pump of *Klebsiella pneumoniae*.";

RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ318073; CAC41008.1; --
 DR InterPro: IPR006143; HLYD.
 DR Pfam: PF00529; HLYD; 1.
 DR SEQUENCE 398 AA; 42426 MW; 57CFA07BE450669 CRC64;

Query Match 33.8%; Score 683; DB 2; Length 398;
 Best Local Similarity 39.3%; Pred. No. 5,6e-30;
 Matches 155; Conservative 80; Mismatches 133; Indels 26; Gaps 7;

11 AAALAAALVALVSSCGKGGDAAGGPGAREAPVGVVYVHPOTVALTYELPERLESL 70
 11 AVLTMLSSGLATGC-DKPAQQAQ-----HMEVGIIVLKSAVPLDITTELPRTSAY 63
 71 RTADVRAQVGGIIQKRLFOEGSYVYAGOPLYOISSTYEAUNLESARQATLQAATLAKAD 130
 64 RIAEVRPQVSGIILKKNFVSGSDIQAGVFLYOIDPATYQSYDAKGLAQALAAANNDQ 123
 131 ADLARKYPLVAEAASVROEYDAVYAKRSAGVKAQAALKSAGININRSRTAPISGF 190
 124 LTVKRYOKLTQYISQDDYDPAVATAQNNNAVVAATVAVETARINLAATYQ-KSPLRSA 182
 191 IGQSK--VSEGTILNAGDTVLATIRQTNPMYVNVVQTSASFVMTLRQIAEGKILAADGV 248
 183 AGSGKSPVTEGALVQNGQTLATVQVDPYVDVDTSSNDPLRKOELADARLKQENGK 242
 249 IAVGKFPDDGTVPYEGKRLFPADPVVNESTGQITLRAVPNDONILMPGLVYRVLMDOVA 308
 243 AKVELVYNDGAKYQSGTLFESDVTYVDQTSITLRAFPNPDHILPGMVRARLEIGI 302
 309 VDNAPVYVPOAVTRGAKD-TWMTVNAQGMPEPREVTVAQOQGTNMTVTSGLKDGDKVYV 366
 303 NPDLALVPOQGVTRTPRGDASVMTVYVGEQKVEVQVTSQAISGKMLVTDLKSGDRVIY 362
 367 EGISTAGTGAKKVT-----KEMASSENOAAA 394
 363 -----TGLOKIKPGVQVAKAEVASDDKQAAA 388

RESULT 31

09BPB7 PRELIMINARY; PRT; 408 AA.
 AC 09BPB7;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Precursor of drug resistance protein.
 GN XF2093.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xylella.
 OX NCBI_TaxID=2371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9a5c;
 RX MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
 RA Alvares R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carter H.,
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
 RA Faciniani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.W., Hohnsbeil J.D., Junqueira M.L., Kemper E.L., Kitaajima J.P.,
 RA Klieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukura A.Y.,
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nhami A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,

RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA da Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuboko M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
 RT "the genome sequence of the plant pathogen Xylella fastidiosa";
 RL Nature 406:151-159(2000).
 DR EMBL: AE004025; AAF84892.1; --
 DR InterPro: IPR006143; HLYD.
 DR Pfam: PF00529; HLYD; 1.
 DR Complete proteome.
 DR SEQUENCE 408 AA; 43636 MW; E08A62B9F3BD8EAB CRC64;

Query Match 33.2%; Score 670; DB 16; Length 408;
 Best Local Similarity 36.8%; Pred. No. 3e-29;
 Matches 146; Conservative 90; Mismatches 149; Indels 12; Gaps 5;

9 MRAAALAAVALVSSCGKGGDAAGGPGAREAPVGVVYVHPOTVALTYELPERLE 68
 17 LRVGLAVYVLLALVATCSK-----SEQ---QMPPEVSIIVAKPQYTPIDLVGRLS 67
 69 SLRTADVRAQVGGIIQKRLFOEGSYVYAGOPLYOISSTYEAUNLESARQATLQAATLAK 128
 68 AYSADVRAVAGILQKRLITTEGTEVKEGQPLQIDPAPQATLLEQGLAAAEATYKN 127
 129 ADLARKYPLVAEAASVROEYDAVYAKRSAGVKAQAALKSAGININRSRTAPIS 188
 128 AKIVADRARLSPQYISRSQVNDVNAEAEARTAAASYEGAAAMQNNRINLVNVTAPIS 187
 189 GFTGQSVSEGTILNAGDTVLATIRQTNPMYVNVVQTSASFVMTLRQIAEGKILAADG 247
 188 GLAQOQVTEGALVSGSSTLITTVQIDPLVNFMSNDELQOLROAQHGSYQSSDN 247
 248 VIAVGKFPDDGTVPYEGKRLFPADPVVNESTGQITLRAVPNDONILMPGLVYRVLMDOV 307
 248 TSTIDVLLSDGSKYFPGIIGIDFSGATVDPSTGVSILRAVLPNDHLLPAPVTFKXNLG 307
 308 AVDNAPVYVPOAVTRGAKD-TWMTVNAQGMPEPREVTVAQOQGTNMTVTSGLKDGDKVYV 366
 308 QRNNTLILPGAGVGRDANSVAVVIDKDGKVIKNVTTIGMKONMILITGLSNGDRVIY 367
 367 EGI-STAGTGAKKVT-----KEMASSENOAAAPOSGVQTA 402
 368 DGLQKVEGAPAKAIPKRPDLISSAHAAAKTPTHTA 404

RESULT 32

09BRY9 PRELIMINARY; PRT; 398 AA.
 AC 09BRY9;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Sma.
 GN SNEA.
 OS Xanthomonas maltophilia (Pseudomonas maltophilia) (Stenotrophomonas
 OS maltophilia).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Stenotrophomonas.
 OX NCBI_TaxID=40324;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=UJA-511;
 RA Li X.-Z., Zhang L., Poole K.;
 RT "A multidrug efflux system, SmeRSABC, of Stenotrophomonas
 maltophilia";
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF173226; AAD51344.1; --
 DR InterPro: IPR006143; HLYD.

DR Pfam; PF00529; HlyD; 1.
SQ SEQUENCE 398 AA; 41888 MW; 2D676251E2C00A9B CRC64;

Query Match 32.8%; Score 662; DB 2; Length 398;

Best Local Similarity 38.0%; Pred. No. 7.9e-29; Mismatches 135; Indels 20; Gaps 6;

Matches 147; Conservative 85; Mismatches 135; Indels 20; Gaps 6;

14 LAALVALVSSCGKGDAAAGQAPGR-EAPA-PVVGVTVPQTVALTELPGRLSESLR 71
14 LLLPLLLALACS-----ADRTDAPAMPEVGVIIASQPLALQQLPGAAVPE 62
14 LLLPLLLALACS-----ADRTDAPAMPEVGVIIASQPLALQQLPGAAVPE 62
72 TADVAQVGGIIQKRLFOEGSVYRAGQPLVQIDSTYEANIESARAQATATQATLAKDA 131
63 ISEVAPQIGLIRQLFTREGQVKAQQLLYQVDPAPYQAAPDTARGQLAQAQEAATLASAP 122
132 DLARKKPLVAEAVROEYDAVATKRSAGEGVKAAQAIKSGINLRSTRTAISGFI 191
123 KAERTPLVSPMAKQADDTATSKQQAQANVIAARAALQARINLDTYRTATFIDRI 182
192 GQSKVSEGLTNAGDTTVALTIRQTNPMVYVNTQSASEVMKLRQIAEGKLLAADGVAV 251
183 GTSSVTAGALVAAAGDDTLTTIGRLDPYLDVTSSTQMLARKLDAAGLVKADGKAQV 242
252 GIKPDDGVYPERKGLLPADPVNVESTGQITRAAVPNDQNLNGLYVRILMDQAVDN 311
243 KVLLEDGSTVYHEGTLFEVGSVDPGTGVKLRVAVIPNDGILLPGMYLKVLPMATDAR 302
312 AFVPAQQAQVTRG--AKDTVMIVNAQGMPEPREVTVAQOQGTWIVTSGLKDGKRVV-EG 368
303 ALLVQKAVNREREGPILRLIDAKDHVERRVSTGVVAGQWQITSGLKAGERVIVSNG 362
369 ISAGITGAKKVPKEMASSENQAAP 395
363 SAVSLGQGVKAVP-----TTAQLAAMP 385

RESULT 33

OBX4LO

PRELIMINARY; PRT; 385 AA.

08X4LO: OBX4LO: PRELIMINARY; PRT; 385 AA.
AC 08X4LO: OBX4LO: PRELIMINARY; PRT; 385 AA.
DT 01-MAR-2002 (TRENBLREL. 20, Created)
DT 01-MAR-2002 (TRENBLREL. 20, Last sequence update)
DT 01-OCT-2002 (TRENBLREL. 22, Last annotation update)
DE Putative membrane protein.
GN YHU OR 24926 OR EC64393.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OC NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RP STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamocis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.,
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7,"
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RP STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kihara S., Shiba T., Hattori M., Shinagawa H.,
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12,"
RL DNA Res. 8:11-22(2001).
DR EMBL; AF005576; AAC58654.1; -.

DR EMBL; AP002565; BAB37816.1; -.
DR InterPro; IPR006143; HlyD.
DR Pfam; PF00529; HlyD; 1.
KW Complete proteome.
SQ SEQUENCE 385 AA; 41220 MW; 6823EF67D6141102 CRC64;

Query Match 32.4%; Score 655; DB 16; Length 385;

Best Local Similarity 38.5%; Pred. No. 1.8e-28; Mismatches 153; Indels 4; Gaps 2;

25 CGKGDAAAGQGPAGREAPAPVVGVTVPQTVALTELPGRLSESLRTADVAQVGGIIQ 84
14 CGAMLTACDDKASANTAMTPEVGVIIASQPLALQQLPGAAVPE 73
85 KRLFOEGSVYRAGQPLVQIDSTYEANIESARAQATATQATLAKADLARKYKVLAAEA 144
74 KRNFIKEDKVNQGSLLVQIDPAPLQALNSAKGLAALSTASNAKRTTFNRQASLKTNY 133
145 VSRQEDYPAATYAKRSABAGVKAQAQAIKSGINLRSTRTAISGFIQSKVSEGLTNA 204
134 VSRQEDYPAATYAKRSABAGVKAQAQAIKSGINLRSTRTAISGFIQSKVSEGLTNA 193
205 GDTTVALTIRQTNPMVYVNTQSASEVMKLRQIAEGKLLAADGVAV 264
194 NQADSLVTVOQLDPIYDLTQSVODFLRKEEVASGQIKOVQSTPVQNLNENKRYST 253
265 GRLFPADPVNVESTGQITRAAVPNDQNLNGLYVRILMDQAVDN 324
254 GTLKFSPDYDETTGSLTIRAFPNPNDGILLPGMYLKVLPMATDAR 313
325 --KDTVMIVNAQGMPEPREVTVAQOQGTWIVTSGLKDGKRVV--TAGITGAKV 380
314 QKATVALILKDDVVKLRLEIASKAIGDQWVTSGLAQDGRVIVSGLRIRPGIKARAI 373
381 TPKEWASSENQ 391
374 SSOENASTESK 384

RESULT 34

OBXCVL1

PRELIMINARY; PRT; 385 AA.

08CVL1: OBXCVL1: PRELIMINARY; PRT; 385 AA.
AC 08CVL1: OBXCVL1: PRELIMINARY; PRT; 385 AA.
DT 01-MAR-2003 (TRENBLREL. 23, Created)
DT 01-MAR-2003 (TRENBLREL. 23, Last sequence update)
DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)
DE Hypothetical lipoprotein yhu precursor.
GN YHU OR C4324.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OC NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RP STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesech P.,
RA Raeko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Moley H.L.T., Donnenberg M.S., Blattner F.R.,
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli,"
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AF016768; AAN82760.1; -.
KW Hypothetical protein; Lipoprotein; Complete proteome.
SQ SEQUENCE 385 AA; 41160 MW; 3DC65B6CCF51CBB8 CRC64;

Query Match 32.4%; Score 655; DB 16; Length 385;

Best Local Similarity 38.5%; Pred. No. 1.8e-28; Mismatches 153; Indels 4; Gaps 2;

25 CGKGDAAAGQGPAGREAPAPVVGVTVPQTVALTELPGRLSESLRTADVAQVGGIIQ 84
14 CGAMLTACDDKASANTAMTPEVGVIIASQPLALQQLPGAAVPE 73
85 KRLFOEGSVYRAGQPLVQIDSTYEANIESARAQATATQATLAKADLARKYKVLAAEA 144
74 KRNFIKEDKVNQGSLLVQIDPAPLQALNSAKGLAALSTASNAKRTTFNRQASLKTNY 133
145 VSRQEDYPAATYAKRSABAGVKAQAQAIKSGINLRSTRTAISGFIQSKVSEGLTNA 204
134 VSRQEDYPAATYAKRSABAGVKAQAQAIKSGINLRSTRTAISGFIQSKVSEGLTNA 193
205 GDTTVALTIRQTNPMVYVNTQSASEVMKLRQIAEGKLLAADGVAV 264
194 NQADSLVTVOQLDPIYDLTQSVODFLRKEEVASGQIKOVQSTPVQNLNENKRYST 253
265 GRLFPADPVNVESTGQITRAAVPNDQNLNGLYVRILMDQAVDN 324
254 GTLKFSPDYDETTGSLTIRAFPNPNDGILLPGMYLKVLPMATDAR 313
325 --KDTVMIVNAQGMPEPREVTVAQOQGTWIVTSGLKDGKRVV--TAGITGAKV 380
314 QKATVALILKDDVVKLRLEIASKAIGDQWVTSGLAQDGRVIVSGLRIRPGIKARAI 373
381 TPKEWASSENQ 391
374 SSOENASTESK 384


```

Db 14 CGAMLTACDYSANAAAMTPEVGVVTLSPGSVNVTLSELPRTVPEYVAEIRPOVGII 73
Qy 85 KRLFOEGSVYRAGOPLYOIDSSTYEANLESARAOATLAKADADLAIEKPLVAEA 144
Db 74 KRNFBEGKVAGOSLSYOIDPAPLOAEINSAKSLAKSLASTASNAKRTIPNQASLIK 133
Qy 145 VSRQETDAVTAKRSAEAGVKAQAIAKSAGINLRSRTIAPISGFIGOSKVSSECTL 204
Db 134 VSRQDYDPAFTQVLEAEANVTAKAAVAEQATINLOYANVTSPITGVSGSKSVTGA 193
Qy 205 GDTIVLTIQRTNPMYVNVVTSASEVMTLRQIAEGKLLADGVYAVIKRDKDGTVP 264
Db 194 NQADSLVTVQRLDPIYDLTVSDPFLMKKEVASGQIKQVSGTPTQJLNLNKRQ 253
Qy 265 GRLLFADPVPVNESTGQITLRAAVPNDQNIIMPGLYVRLAMDQVAVDPAFVPO 324
Db 254 GTLFPSDPTVDETTGSLTIRAFIPNPNDLLPGMVTALVDEGRQVNLVLPQSGV 313
Qy 325 --KDTMTVNAOGMEPREVTVAOQGTNMTVTSGLKDKGVVVGIS--TAGITGAK 380
Db 314 QGKATALLDQDQVQLEIEBASKAIGQVMTVSGLOQGRVIVSGLQRIKRAAIS 373
Qy 381 TPKEWASSENQ 391
Db 374 SSQENASTESK 384

RESULT 35
Q9KRW5 PRELIMINARY; PRT; 382 AA.
AC Q9KRW5;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE Periplasmic linker protein.
GN TIGD.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2011535; PubMed=10648517;
RA Mosqueda G., Ramos J.L.;
RA Mosqueda G., Ramos J.L.;
RT "A set of genes encoding a second toluene efflux system in Pseudomonas
RT putida Dof-T1 is linked to the tod genes for toluene metabolism.";
RL J. Bacteriol. 182:937-943(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Mosqueda-cano G.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL, Y19106; CAB72258.1; -.
DR InterPro; IPR006143; HlyD.
DR Pfam; PF00529; HlyD; 1.
SQ SEQUENCE 382 AA; 41584 MW; EE2C751979592757 CRC64;

Query Match 32.4%; Score 654; DB 2; Length 382;
Best Local Similarity 39.0%; Pred. No. 2,1e-26;
Matches 158; Conservative 70; Mismatches 143; Indels 34; Gaps 10;

```

```

Db 164 PITGRIRSAVTREGALVTNQAQELATVTOQDPIYVNVNQPITRLGKRALESGRLO 223
Qy 245 ADVGIAVGIFPDGTVPPEKGRLLFADPVNNESTGQITLAAVNDQNIIMPGLYVRL 304
Db 224 GDNQAVSLTLDSDTPYPLGVLFSEVSVA PSTGSVTLAEFPNPHKLLPGFVAL 283
Qy 305 DOVAVDPAFVVPPOAVTRGAK--DTWMTVNAOGMEPREVTVAOQGTNMTVTSGLK 362
Db 284 NEGGQQAAILVPHQAVGRDARGVPTVWVVPDVTVEREVQITQVGNALLAGINDGE 343
Qy 363 KVVVEGISA--GITGAKVTPKEWASSENQAAAPQSGVOTASEA 405
Db 344 RVITGEVQLARSGI----VKP---VAKKVKLMSEFGSQVQAQA 381

RESULT 36
Q8P876 PRELIMINARY; PRT; 436 AA.
AC Q8P876;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE Multidrug resistance protein.
GN ACRA OR XC2367.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Camarvan F., Cardozo J., Chamberggo F., Cláudia L.P.,
RA Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.O.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A.M., Sena J.A.D., Silva C. de Souza R.F.,
RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.";
RL Nature 417:459-463(2002).
DR EMBL, AE012345; AA041645.1; -.
DR InterPro; IPR006143; HlyD.
DR Pfam; PF00529; HlyD; 1.
KM Complete proteome.
SQ SEQUENCE 436 AA; 44879 MW; EB45096A5C9CE011 CRC64;

Query Match 32.1%; Score 647.5; DB 16; Length 436;
Best Local Similarity 35.0%; Pred. No. 5,6e-28;
Matches 150; Conservative 83; Mismatches 154; Indels 41; Gaps 6;

```

QY 188 SGFIGQSVSEGLTLNAGDTVLATIRQTNPMYVNTQSASEVWKLRLQIAEGKL-LAAD 246
 Db 176 TGRAGIQVTEGALVGSSEATLITVDQIDPLVYFAMSSSELAALRQAGSGNVQJSGD 235
 QY 247 GVIAVGIRFDDGTVPKGRLLFPADVNESTGQITLRAAIPNDONILMPGLYRVLMDQ 306
 Db 236 GKSTINVELNGTOYHPHGTLDVSAVTDPSTGAVSLRATLFPNPMALLPGAFTVFKASL 295
 QY 307 VAVDNAFVFPQOAVTRGAKDT-VMIVNAQGMREPREVTVAQOQGNMTVITSGLKDGDKV 365
 Db 296 GQRNNAIVFPQALORDATGAYALVIGEDGKVRKGLTVDDQKQKQMTVTHAGMAGDQVI 355
 QY 366 VEGISIA-----GITGAKKVTPEKMASSENO-----AAPQ 396
 Db 356 VDGVOKAKGEPKAGVPMDPNKPAGQGPAGAPGAAPAGAGTGGAGPAGAKADAAAPA 415
 QY 397 SGVOTASE 404
 Db 416 AGEQPKSD 423

RESULT 37

Q93E20 PRELIMINARY; PRT; 396 AA.
 AC Q93E20;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
 DE AdeA membrane fusion protein.
 GN ADEA.
 OS Acinetobacter baumannii.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Moraxellaceae; Acinetobacter.
 CX NCBI_TaxID=470;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BM4454;
 RA Magnet S., Courvalin P., Lambert T.;
 RT "Characterization of a RND type efflux pump involved in aminoglycoside
 RL resistance in Acinetobacter baumannii clinical isolate."
 DR Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF370885; AAL14439.1; --
 DR InterPro; IPR006143; HLYD.
 DR Pfam; PF00529; HLYD; 1.
 SQ SEQUENCE 396 AA; 43376 MW; CABFCD53472F80B CRC64;

Query Match 32.0%; Score 647; DB 2; Length 396;
 Best Local Similarity 38.3%; Pred. No. 5.3e-28;
 Matches 152; Conservative 75; Mismatches 156; Indels 14; Gaps 5;

QY 14 LAAAVALTSSCGKGDAAQGGPAGREAPAVGVTVHPQTVALTVELPGRLESLRTA 73
 Db 9 LFLSLGLLIGCC-----DSKEVQAQAEPPRAKVSVLISIQPOSVFSENLPKRVIAFRTA 61
 QY 74 DVRAQVGGIIQKRLFOEGSYVRAGPOLYQIDSTYEANLESARQALTAQATLAKADL 133
 Db 62 EIRPQVGGIIEKVLFKQSEVVRAGQALYKINSETFEADVNSVRASLNAEAEVARTKQVL 121
 QY 134 ARKRLVLAAEAVSRQEDYDAATYAKSAEAGVKAADAKSKAGININRSRITRPISGFIGQ 193
 Db 122 ERYEELLSNMAISKQVSNAAQYQALADAVQAQMALARONLNIQYATVRAPISGRIGQ 181
 QY 194 SKVSEGLTLNAGDTVLATIRQTNPMYVNTQSASEVWKLRLQIAEGKL-LAADGVIAVG 253
 Db 182 SFVTGALVGGDDTMTATIQIDKIVYDVQKQSVSEYERLQALOSGE-LSANSKIVRI 240
 QY 254 KFDGTVPKGRLLFPADVNESTGQITLRAAVENDONILMPGLYRVLMDQVADNAF 313
 Db 241 TNSHQPNVTVAKMLFEDINVDPEIGDVTFRLEVNNTERKLLPGMYVAVNIDRASIPQAL 300
 QY 314 VVPOQAVTR--GAKTVMIVNAQGMREPREVTVAQOQGNMTVITSGLKDGDKRVVVGISI 371
 Db 301 LVPAQAIQRNISGEPQVYVINAQGTAEIRPIEIGQYEQPYIANKGLKVGRRVVEGIE- 359

QY 372 AGITGAKKVTPEKMASS--ENQAAAPQSGVOTASEAK 406
 Db 360 -RIKPNQKLALAAWKAPAVANHAASVETKTSIAEGAQ 395

RESULT 38

O31099 PRELIMINARY; PRT; 382 AA.
 ID O31099;
 AC O31099;
 DT 01-JAN-1998 (Tremblrel. 05, Created)
 DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
 DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
 DE Periplasmic linker protein.
 GN SRPA.
 OS Pseudomonas putida.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 CX NCBI_TaxID=303;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S12;
 RX MEDLINE=98079031; PubMed=9417051.
 RA Kieboom J., Dennis J.U., de Bont J.A., Zylstra G.J.;
 RT "Identification and molecular characterization of an efflux pump
 RL involved in Pseudomonas putida S12 solvent tolerance."
 RN J. Biol. Chem. 273:85-91(1998).
 RP SEQUENCE FROM N.A.
 RC STRAIN=S12;
 RX MEDLINE=99069360; PubMed=9852029.
 RA Kieboom J., Dennis J.U., Zylstra G.J., de Bont J.A.;
 RT "Active efflux of organic solvents by Pseudomonas putida S12 is
 RL induced by solvents."
 RN J. Bacteriol. 180:6769-6772(1998).
 RP [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S12;
 RA Kieboom J., Dennis J.U., Zylstra G.J., de Bont J.A.M.;
 RL Submitted (Feb-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF029405; AAD12178.1; --
 DR InterPro; IPR006143; HLYD.
 DR Pfam; PF00529; HLYD; 1.
 SQ SEQUENCE 382 AA; 41429 MW; 8B7AB0A2DC4C440A CRC64;

Query Match 31.7%; Score 640; DB 2; Length 382;
 Best Local Similarity 36.0%; Pred. No. 1.2e-27;
 Matches 143; Conservative 84; Mismatches 138; Indels 32; Gaps 8;

QY 7 KAMRAALAAVAVLTSSCGKGDAAQGGPAGREAPAVGVTVHPQTVALTVELPGR 66
 Db 8 RALAVITLTA--LMLISGCE-----KEQVSATPPPVGVTVRAQALTLTTDLPGR 58
 QY 67 LESLRTADVRAQVGGIIQKRLFOEGSYVRAGPOLYQIDSTYEANLESARQALTAQATL 126
 Db 59 TSARFVVEVPQVSGIIQKRSFVEGAEVKLGOQLYQIDPRYEQRLRAEAMRTSAQ-NL 117
 QY 127 AKADADLARYKPLVAAAVSRQEDYDAATYAKSAEAGVKAADAKSKAGININRSRITAP 186
 Db 118 AR-----RIETLTKTAVSKQQTDDLAAMKQAEAYQVAR-----IDVQTRVUSP 164
 QY 187 ISGFIGQSVSEGLTLNAGDTVLATIRQTNPMYVNTQSASEVWKLRLQIAEGKL-LAA 245
 Db 165 ISGRIGSTVTEGALVTNGAQSLATVQIDPIVDVOTPIITKLGLQKALLESGRLOXTG 224
 QY 246 DGVIAVGIRFDDGTVPKGRLLFPADVNESTGQITLRAAVENDONILMPGLYRVLMD 305
 Db 225 ENQAEVSLTLDDGSAAYLPGLTKFSEVSVDPTGSVTLRAEPNPKLLPGMFAHLK 284
 QY 306 QVAVDNAFVVPQOAVTRGAK--DTVMIVNAQGMREPREVTVAQOQGNMTVITSGLKDGDK 363
 Db 285 EGVQMAILVFPQAIISRDTRGVPSVWVVKADNTVESREIOTLRTKVGAMALISNGVTEGER 344

Qy 364 VVEGIS-----IAGITGAKKVTPEKMASSENQAAA 394
 Db 345 IITEGVORVRSGLAVNAVEAKVNLVDPGFAATTEASA 381

RESULT 39

Q8PJN2 PRELIMINARY; PRT; 435 AA.
 ID Q8PJN2
 AC Q8PJN2;
 DT 01-OCT-2002 (Tremblrel. 22, Created)
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
 DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
 DE Multidrug resistance protein.
 GN ACRA OR XAC2498.
 OS Xanthomonas axonopodis (pv. citri).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xanthomonas.
 CX NCBI_TaxID=92829;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=306 / ATCC 13902 / XV 101;
 RX MEDLINE=22022145; PubMed=12024217;
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
 RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.B.A.,
 RA Caramante G., Cammaran F., Cardozo J., Chamberg F., Ciapina L.P.,
 RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
 RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
 RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
 RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
 RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
 RA Mota L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
 RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
 RA Spinoza L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
 RA Trindade dos Santos W., Truffi D., Tsai S.M., White F.F.,
 RA Setubal J.C., Kitajima M.J.P.;
 RT "Comparison of the genomes of two Xanthomonas pathogens with differing
 RT host specificities";
 RT host specificities";
 RL Nature 417:459-463(2002).
 DR EMBL; AE011888; AA037349.1; -;
 DR InterPro; IPR006143; HLYD.
 DR Pfam; PF00529; HLYD; 1.
 KW Complete proteome.
 SQ SEQUENCE 435 AA; 4959 MW; CEB07451276FEB6D CRC64;

Query Match 31.5%; Score 636; DB 16; Length 435;
 Best Local Similarity 36.1%; Pred. No. 2.4e-27;
 Matches 151; Conservative 79; Mismatches 162; Indels 26; Gaps 6;
 Qy 9 MRAAALAAVAVLVSSCGKGDAAGOPAGREAPAPVGVVTVHQTVALTVTELGR 67
 Db 5 LRTGLCALITVAALAAACK-----PDQQAAPPEPVSVLEMKPTTLEEDLVGR 55
 Qy 68 ESLRTADVRAQVGGIIQKRLFOEGSYVRAGOPLYOIDSSTYEANLESARAQATAQT 127
 Db 56 SAYASADYRAVDEVLKRLYTBEGANTYEGPFLQIDPSQKATLLQAQGLAAEAATYT 115
 Qy 128 KADADLARYKPLVAEAVSRQEDYDAVYAKTSABAGVKAQAQAIKAGININRSRTAP 187
 Db 116 NAKIAARARASLAPQOYVSRAIDTAEANERSGANNQOARGAVEAARIQGFATVSP 175
 Qy 188 SGFIGSKVSEGLTLNAGDTTVLATITROTNPVYVNTOSASEVMKLRQIAEGTL-LA 246
 Db 176 TGRAGIQRVTEGALVAGEATLLTTVDQIDPLVNPAMSSSELAALQAQSSGNVQLSG 235
 Qy 247 GVIAVGIKFDGTVYPEKGRLLFADPVVNESTGOITLRAAVNDONILMPGLYRVLM 306
 Db 236 GKSSINVELNGTQYQHPGTLIDVASAVYDPTGAVSLRALTPNPGSLIPRAFTFRAS 295
 Qy 307 VAVDNAFVFPQAAVTRGAKDT-VMI VNAQGMREPVTVAAQOGTNNI VTSGLKDG 365
 Db 296 GORNNAYLLPQAAVORBATGAYALVLGKDGKVRKKNLTVDQQKQGWIVTGMTPTGD 355

Qy 366 VEGISIAITGAKKVTPEK-----ASSENQAAAPOSQVOTASEAKTASEA 411
 Db 356 VDGVRKAKEGQPAKGVF--WDPNKPAQOGGGAAGAAPAAQGGDAQAAPADRADAA 411

RESULT 40

Q93PUS PRELIMINARY; PRT; 391 AA.
 ID Q93PUS
 AC Q93PUS;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
 DE Periplasmic efflux pump component Tegg.
 GN TTGG.
 OS Pseudomonas putida.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 CX NCBI_TaxID=303;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DOT-TIE;
 RX MEDLINE=21289082; PubMed=11395460;
 RA Rojas A., Duque E., Mosqueda G., Golden G., Hurtado A., Ramos J.L.,
 RA Segura A.;
 RT "Three Efflux Pumps Are Required To Provide Efficient Tolerance to
 RT Toluene in Pseudomonas putida DOT-TIE";
 RL J. Bacteriol. 183:3967-3973(2001).
 DR EMBL; AF299253; AA069563.1; -;
 DR InterPro; IPR006143; HLYD.
 DR Pfam; PF00529; HLYD; 1.
 SQ SEQUENCE 391 AA; 4258 MW; E2362606D71E52A1 CRC64;

Query Match 31.3%; Score 631; DB 2; Length 391;
 Best Local Similarity 35.5%; Pred. No. 3.9e-27;
 Matches 141; Conservative 84; Mismatches 140; Indels 32; Gaps 8;
 Qy 7 KARAALAAVAVLVSSCGKGDAAGOPAGREAPAPVGVVTVHQTVALTVTELGR 66
 Db 17 RALRVPLTA--LMLISGCE-----KEQVSATTPPDVGVTVVAAQALTLTTDLPGR 67
 Qy 67 LESLRTADVRAQVGGIIQKRLFOEGSYVRAGOPLYOIDSSTYEANLESARAQATAQT 126
 Db 68 TSAFRVAEVRAPQVSGILQKRSFVEGAHVKLQGLQYQIDPRTYEQLRRAEANRRTSAQ-NL 126
 Qy 127 AKADADLARYKPLVAEAVSRQEDYDAVYAKTSABAGVKAQAQAIKAGININRSRTAP 186
 Db 127 AR-----RYETLLKTKAVSKQOYDDALAAMKQAEADYQVAR-----IDVYTRVLS 173
 Qy 187 ISGFIGSKVSEGLTLNAGDTTVLATITROTNPVYVNTOSASEVMKLRQIAEGTL-LA 245
 Db 174 ISGRIGSTVTEGALVNGAQSILATVTDPLIYVDTPQITKLGLHKLBSGRLOKTC 233
 Qy 246 DGVIANGIKFDGTVYPEKGRLLFADPVVNESTGOITLRAAVNDONILMPGLYRVLM 305
 Db 234 ENQAEVSLTDDSAVDLPGLTKFSEVSCDPTGVSVTLRAEPNPNKLLPGMVFHALL 293
 Qy 306 QVAVDNAFVFPQAAVTRGAK--DTVMIVNAQGMREPVTVAAQOGTNNI VTSGLKDG 363
 Db 294 EGVQNAAILPQOALISHDTPGVPSVWWMKADNVESEIOTLRTVGNAMLISNGVTGER 353
 Qy 364 VVEGIS-----IAGITGAKKVTPEKMASSENQAAA 394
 Db 354 IITEGVORVRSGLAVNAVEAKVNLVDPGFAATTEASA 390

Search completed: September 8, 2003, 14:02:11
 Job time : 45 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2003, 13:59:31 / Search time 21 Seconds
(without alignments)
1886.738 Million cell updates/sec

Title: US-09-889-756A-2

Perfect score: 2019
Sequence: 1 MAFYAFKRAMRAALAAVAL.....AAPQSGVQTASEAKTASEAE 412

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: PIR_76: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1993	98.7	412	2	H81825
2	1989	98.5	412	2	H81051
3	1278	63.3	271	2	S42418
4	770	38.1	388	2	H87349
5	762.5	37.8	397	2	D90693
6	762.5	37.8	397	2	H85543
7	762.5	37.8	397	2	A36938
8	756	37.4	397	2	AG0561
9	753	37.3	383	2	G39629
10	750.5	37.2	385	2	G65119
11	749.5	37.1	385	2	A91146
12	749.5	37.1	385	2	H85991
13	726.5	36.0	395	2	AD0380
14	723	35.8	404	2	AE2950
15	723	35.8	427	2	F98332
16	699.5	34.6	387	2	A83072
17	697.5	34.5	373	2	G90861
18	697.5	34.5	373	2	G90861
19	670	33.2	408	2	B82600
20	656	32.5	385	2	S47723
21	655	32.4	385	2	A91178
22	655	32.4	385	2	B86024
23	628.5	31.1	386	2	AD0423
24	628	31.1	396	2	B83393
25	595	29.5	400	2	AE2889
26	595	29.5	410	2	B97665
27	537	26.6	395	2	AH3455
28	482	23.9	388	2	G95375
29	460	22.8	414	2	E83335

30	459.5	22.8	373	2	G96007	probable acriflavi
31	446	22.1	414	2	T30829	hypothetical prote
32	442.5	21.9	396	2	AG2881	HYD family secret
33	442.5	21.9	396	2	G97657	hypothetical prote
34	432	21.4	464	2	E85834	probable membrane
35	432	21.4	464	2	A64974	hypothetical prote
36	431	21.3	413	2	AC0771	probable efflux sy
37	431	21.3	464	2	B90989	probable membrane
38	430	21.3	426	2	A83330	probable RND efflu
39	426.5	21.1	451	2	A13623	acriflavin resista
40	425.5	21.1	444	2	AH0346	putative HYD fami
41	424.5	21.0	367	2	E81379	probable membrane
42	403.5	20.0	425	2	AH0122	probable drug effl
43	394	19.5	43	2	A83466	probable RND efflu
44	387	19.2	411	2	A82566	membrane fusion pr
45	356	17.6	425	2	AC2959	HYD family secret

ALIGNMENTS

RESULT 1									
H81825									
membrane fusion protein NMA1970 [imported] - Neisseria meningitidis (strain Z2491 serogr									
C/Species: Neisseria meningitidis									
C/Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001									
C/Accession: H81825									
R/Parikh, J.; Achman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morell									
; Holroyd, S.; Jørgels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,									
Nature 404, 502-506, 2000									
A/Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.									
A/Reference number: AB1775; MUID:20222556; PMID:10761919									
A/Accession: H81825									
A/Status: preliminary									
A/Molecule type: DNA									
A/Residues: 1-412 <PAR>									
A/Cross-references: GB:AL162757; GB:AL157959; NID:97380371; PIDN:CAB85190.1; PID:97380604.									
A/Experimental source: serogroup A, strain Z2491									
C/Genetics:									
A/Gene: mtrC; NMA1970									
Query Match									
Best Local Similarity 98.7%; Score 1993; DB 2; Length 412;									
Matches 406; Conservative 2; Mismatches 4; Indels 0; Gaps 0;									
QY	1	MAFYAFKRAMRAALAAVALVLS	CGKGGDAAAGGQAGREAPVGVVTHPQVALT	60					
DB	1	MAFYAFKRAMRAALAAVALVLS	CGKGGDAAAGGQAGREAPVGVVTHPQVALT	60					
QY	61	VELPGRLESIRTAADVRAOVGGIIQKRLFORGSVVRAGQPIYQIDSSTYEANLESARQOLA	120						
DB	61	VELPGRLESIRTAADVRAOVGGIIQKRLFORGSVVRAGQPIYQIDSSTYEANLESARQOLA	120						
QY	121	TAQATLAKADADADARVYPLVAEAVEROEYDAVTAKRSAEAGVKAQAQAIKSAGINLNR	180						
DB	121	TAQATLAKADADADARVYPLVAEAVEROEYDAVTAKRSAEAGVKAQAQAIKSAGINLNR	180						
QY	181	SRTTAPISGFTGSKVSEGLTLNAGDTTVLATIRQTNPMVYVNTQSASEVMKLRQIAEG	240						
DB	181	SRTTAPISGFTGSKVSEGLTLNAGDTTVLATIRQTNPMVYVNTQSASEVMKLRQIAEG	240						
QY	241	KLLAADGVAVGKPDGTVYPERKRLFPADPVVNSGQITLRAVPNDQNLIMGLVY	300						
DB	241	KLLAADGVAVGKPDGTVYPERKRLFPADPVVNSGQITLRAVPNDQNLIMGLVY	300						
QY	301	RVLMDQVAVDNAPVFPQAVTRGAKDTVMIVNAQGMREPREVVAQOQGNMIVTSGLKD	360						
DB	301	RVLMDQVAVDNAPVFPQAVTRGAKDTVMIVNAQGMREPREVVAQOQGNMIVTSGLKD	360						
QY	361	GDKRVVEGISIAGITAKKVTYPRKMASSENOAAPOSGVQTASEAKTASEAE	412						
DB	361	GDKRVVEGISIAGITAKKVTYPRKMASSENOAAPOADVQTASEAKTASEAE	412						

RESULT 2
F81051
membrane fusion protein NMB1716 [imported] - Neisseria meningitidis (strain MC58 serogroup C)
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jun-2001
C:Accession: F81051
R:Jettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; Ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M. Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Vignani, A.
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: AB1000; MUID:20175755; PMID:10710307
A:Accession: F81051
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-412 <TEXT>
A:Cross-references: GB:AE002521; GB:AE002098; NID:g7226962; PIDN:AA42063.1; PID:g722697
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB1716

Query Match 98.5%; Score 1989; DB 2; Length 412;
Best Local Similarity 98.5%; Pred. No. 6,4e-114;
Matches 406; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 MAFYAFKMRRAAALAAVALVSSCGKGDAAGGAPAGREAPAPVGVVTHPQTVALT 60
DB 1 MAFYAFKMRRAAALAAVALVSSCGKGDAAGGAPAGREAPAPVGVVTHPQTVALT 60
QY 61 VELPGRLESRTADVRAOVGGIIQKRLFOEGSYVRACOPLYQIDSTYEANLESARAQLA 120
DB 61 VELPGRLESRTADVRAOVGGIIQKRLFOEGSYVRACOPLYQIDSTYEANLESARAQLA 120
QY 121 TAOATLAKADADLARYPLVAEAVSRQBYDAATVAKSAEAGVKAQAQAIKSAGINLR 180
DB 121 TAOATLAKADADLARYPLVAEAVSRQBYDAATVAKSAEAGVKAQAQAIKSAGINLR 180
QY 181 SRTAPISGFIGQSKVSEGTLLNAGDTTLATIRQTNPMYVNTQSAEVMKLRQIAEG 240
DB 181 SRTAPISGFIGQSKVSEGTLLNAGDTTLATIRQTNPMYVNTQSAEVMKLRQIAEG 240
QY 241 KLLAADGVIAVGIRKDDGTVPYPERKRLFPADPVNVESTGQITLRAAVNDQNIIMPGLYV 300
DB 241 KLLAADGVIAVGIRKDDGTVPYPERKRLFPADPVNVESTGQITLRAAVNDQNIIMPGLYV 300
QY 301 RVLMDQVAVDNAFVVPQOAVTRGAKDTWIVNAOCGMPEPREVTVAOQCGTWTIVTSGLKD 360
DB 301 RVLMDQVAVDNAFVVPQOAVTRGAKDTWIVNAOCGMPEPREVTVAOQCGTWTIVTSGLKD 360
QY 361 GDKVNVGISTAGTGAKKVTPKEMWASSENQAAAFOSQVOTRASEAKTASEAK 412
DB 361 GDKVNVGISTAGTGAKKVTPKEMWASSENQAAAFOSQVOTRASEAKTASEAK 412

RESULT 3
S42418
probable antibiotic resistance protein mtrC - Neisseria gonorrhoeae (strain CH95) (fragm
C:Species: Neisseria gonorrhoeae
C:Date: 07-May-1998 #sequence_revision 15-May-1998 #text_change 08-Oct-1999
C:Accession: S42418; S40252
R:Pan, W.; Spratt, B.G.
Mol. Microbiol. 11, 769-775, 1994
A:Title: Regulation of the permeability of the gonococcal cell envelope by the mtr sys
A:Reference number: S42417; MUID:94254732; PMID:8196548
A:Accession: S42418
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-271 <PAN>
A:Cross-references: EMBL:Z25796; NID:g438190; PIDN:CAAB1046.1; PID:g438192
A:Experimental source: CH95
C:Genetics:

A:Gene: mtrC
C:Superfamily: 1:ipoy1/biotin-binding homology
F:68-111,183-211/Domain: 1:ipoy1/biotin-binding homology #status atypical <LPB>
Query Match 63.3%; Score 1278; DB 2; Length 271;
Best Local Similarity 96.7%; Pred. No. 7.1e-71;
Matches 262; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAFYAFKMRRAAALAAVALVSSCGKGDAAGGAPAGREAPAPVGVVTHPQTVALT 60
DB 1 MAFYAFKMRRAAALAAVALVSSCGKGDAAGGAPAGREAPAPVGVVTHPQTVALT 60
QY 61 VELPGRLESRTADVRAOVGGIIQKRLFOEGSYVRACOPLYQIDSTYEANLESARAQLA 120
DB 61 VELPGRLESRTADVRAOVGGIIQKRLFOEGSYVRACOPLYQIDSTYEANLESARAQLA 120
QY 121 TAOATLAKADADLARYPLVAEAVSRQBYDAATVAKSAEAGVKAQAQAIKSAGINLR 180
DB 121 TAOATLAKADADLARYPLVAEAVSRQBYDAATVAKSAEAGVKAQAQAIKSAGINLR 180
QY 181 SRTAPISGFIGQSKVSEGTLLNAGDTTLATIRQTNPMYVNTQSAEVMKLRQIAEG 240
DB 181 SRTAPISGFIGQSKVSEGTLLNAGDTTLATIRQTNPMYVNTQSAEVMKLRQIAEG 240
QY 241 KLLAADGVIAVGIRKDDGTVPYPERKRLFPAD 271
DB 241 KLLAADGVIAVGIRKDDGTVPYPERKRLFPAD 271

RESULT 4
E87349
HlyD family secretion protein [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: E87349
R:Nielsen, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.A.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; Ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: AB7249; MUID:21173698; PMID:11259647
A:Accession: E87349
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-388 <STO>
A:Cross-references: GB:AE005673; NID:g13422055; PIDN:AAK22793.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC0808

Query Match 38.1%; Score 770; DB 2; Length 388;
Best Local Similarity 46.2%; Pred. No. 9.5e-40;
Matches 180; Conservative 57; Mismatches 11; Indels 22; Gaps 9;

QY 16 AAVALVSSCG-KGGDAAGGAPAGREAPAPVGVVTHPQTVALTYELPGRLESRTAD 74
DB 15 AAVALVSSCG-KGGDAAGGAPAGREAPAPVGVVTHPQTVALTYELPGRLESRTAD 74
QY 75 VRAOVGGIIQKRLFOEGSYVRACOPLYQIDSTYEANLESARAQLA-ATAOATLAKADAD 132
DB 70 VRAOVGGIIQKRLFOEGSYVRACOPLYQIDSTYEANLESARAQLA-ATAOATLAKADAD 132
QY 133 LARYPLVAEAVSRQBYDAATVAKSAEAGVKAQAQAIKSAGINLRSTAPISGFIG 192
DB 130 --RKAALVERGAVSRQNDQDAQAALQTAAVGVQKALASARINLVARTATISGFIG 187
QY 193 QSKVSEGTLLNAGDTTLATIRQTNPMYVNTQSAEVMKLRQIAEGKLLAADGVIAVG 252
DB 188 KSSVTAGALVTAQAATLATAVQDSKYVVDLTQTSALKLQAFASGK-VGRSGSAQVT 246
QY 253 IKPDDGTVPYPERKRLFPADPVNVESTGQITLRAAVNDQNIIMPGLYVRLMDQVAVDNA 312
DB 247 LKLEDGISTYPIGRLESRTADVRAOVGGIIQKRLFOEGSYVRACOPLYQIDSTYEANLESARAQLA 306

Qy 313 FVVPQOATVTRAKD--TWTIVNAGGMEPREVTVAOQOGTWTIVTSGLKDGKXVVEGIS 370
Db 307 ILIPGTAVNRDPRKGGATVWLVAOK-GPEPRVPTLQOTVGDKWLVTSGINAGDKVITVEGL- 364
Qy 371 IAGITGAKKVTPEKVASSENQAAAPQSGVQ 400
Db 365 -----MKVPRGAPITKAVPRAGAAPAAQAQ 387

RESULT 5

d90693
acridine efflux pump [imported] - Escherichia coli (strain O157:H7, substrain RIMD 05099
C/Species: Escherichia coli
C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C/Accession: D90693
R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A/Reference number: A99629; MUID:21156231; PMID:11258796
A/Accession: D90693
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-397 <HAY>
A/Cross-references: GB:BA000007; PIDN:BA833939.1; PID:G13359973; GSPDB:GN00154
A/Experimental source: strain O157:H7, substrain RIMD 0509952
C/Genetics:
A/Gene: EC60516

Query Match 37.8%; Score 762.5; DB 2; Length 397;
Best Local Similarity 42.2%; Pred. No. 2.8e-39;
Matches 168; Conservative 76; Mismatches 129; Indels 25; Gaps 6;

Qy 11 AALAAAVAVLSSCGKGGDAAGGAPAGRAPVVGVTVHTVQVATVTELPGRLESL 70
Db 11 AVVLMSSGLATGTC-DDKQAQGGQ-----QMPAVGVTVKTEPLQITTELPGRISAY 63
Qy 71 RTADVRAQVGGIIOKRLFOESGYRAGOPLYOIDSSTEYANLESARQALATQAATLAKAD 130
Db 64 RIAVRPQVSGIILKRNFKESDIEAGVSLVQIDPATYQATYDSKGLAKAQAANIQAQ 123
Qy 131 ADLARKYPLVAEAVSRQEDYAAVTAKRSAGVKAQAATKSAGININRSRTAPISGF 190
Db 124 LTVNRVQKLTGTQYISKQEDQALADAOANAATAAQAAYETARINLAATKVTSPISGR 183
Qy 191 IGGKSVSGTLLNAGDTTVAITROTNPVYNTQSASEWKLKROIAEGTLAADGVIA 250
Db 184 IGSNVTGALVQNGQATATATVQQLDPIYDVTOSSNDPLRLKQELANGTLKQENGKAK 243
Qy 251 VGIFEDGTVPYPEKGRLLFADPVVNESTGOITLRAAVNDONILMPGLYRVLMDOVAVD 310
Db 244 VSLTISGIRFPQGTLEFSDVTVDTGTSITLRAIFNPHTLLPGFVARAREGLNP 303
Qy 311 NAFVVPQOATVTRAKD--TWTIVNAGGMEPREVTVAOQOGTWTIVTSGLKDGKXVVEG 368
Db 304 NALIVPQOGVTRTPRGATVAVGADKVEETRPVSAQIDKMLVTEGLAGDRVVISG 363
Qy 369 ISLAGITGAKKVT-----KEMASSENQAAAPQSGVQ 400
Db 364 L-----QKVRGQVQKAEVTADNNQAA--SGAQ 391

RESULT 6

H85543
acridine efflux pump [imported] - Escherichia coli (strain O157:H7, substrain EDJ933)
C/Species: Escherichia coli
C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C/Accession: H85543
R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimaienta, E.; Potamouisis, K.; Apodaca,
Nature 409, 529-533, 2001
A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A/Reference number: A85480; MUID:21074935; PMID:11206551

A/Accession: H85543
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-397 <STD>
A/Cross-references: GB:AE005174; NID:G12513330; PIDN:AA654812.1; GSPDB:GN00145; UMGCP:205
A/Experimental source: strain O157:H7, substrain EDJ933
C/Genetics:
A/Gene: acra

Query Match 37.8%; Score 762.5; DB 2; Length 397;
Best Local Similarity 42.2%; Pred. No. 2.8e-39;
Matches 168; Conservative 76; Mismatches 129; Indels 25; Gaps 6;

Qy 11 AALAAAVAVLSSCGKGGDAAGGAPAGRAPVVGVTVHTVQVATVTELPGRLESL 70
Db 11 AVVLMSSGLATGTC-DDKQAQGGQ-----QMPAVGVTVKTEPLQITTELPGRISAY 63
Qy 71 RTADVRAQVGGIIOKRLFOESGYRAGOPLYOIDSSTEYANLESARQALATQAATLAKAD 130
Db 64 RIAVRPQVSGIILKRNFKESDIEAGVSLVQIDPATYQATYDSKGLAKAQAANIQAQ 123
Qy 131 ADLARKYPLVAEAVSRQEDYAAVTAKRSAGVKAQAATKSAGININRSRTAPISGF 190
Db 124 LTVNRVQKLTGTQYISKQEDQALADAOANAATAAQAAYETARINLAATKVTSPISGR 183
Qy 191 IGGKSVSGTLLNAGDTTVAITROTNPVYNTQSASEWKLKROIAEGTLAADGVIA 250
Db 184 IGSNVTGALVQNGQATATATVQQLDPIYDVTOSSNDPLRLKQELANGTLKQENGKAK 243
Qy 251 VGIFEDGTVPYPEKGRLLFADPVVNESTGOITLRAAVNDONILMPGLYRVLMDOVAVD 310
Db 244 VSLTISGIRFPQGTLEFSDVTVDTGTSITLRAIFNPHTLLPGFVARAREGLNP 303
Qy 311 NAFVVPQOATVTRAKD--TWTIVNAGGMEPREVTVAOQOGTWTIVTSGLKDGKXVVEG 368
Db 304 NALIVPQOGVTRTPRGATVAVGADKVEETRPVSAQIDKMLVTEGLAGDRVVISG 363
Qy 369 ISLAGITGAKKVT-----KEMASSENQAAAPQSGVQ 400
Db 364 L-----QKVRGQVQKAEVTADNNQAA--SGAQ 391

RESULT 7

A36938
acriflavin resistance protein acra precursor - Escherichia coli (strain K-12)
C/Species: Escherichia coli
C/Date: 11-Nov-1994 #sequence_revision 11-Nov-1994 #text_change 01-Mar-2002
C/Accession: A36938; F64776
R/Ma, D.; Cook, D.N.; Alberti, M.; Pon, N.G.; Nikaido, H.; Hearst, J.E.
J. Bacteriol. 175, 6299-6313, 1993
A/Title: Molecular cloning and characterization of acra and acrc genes of Escherichia co
A/Reference number: A36938; MUID:94012493; PMID:8407802
A/Accession: A36938
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-397 <MA>
A/Cross-references: GB:U00734; NID:G392830; PIDN:AA67134.1; PID:G5523310
R/Baltner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A/Title: The complete genome sequence of Escherichia coli K-12.
A/Reference number: A64720; MUID:97426617; PMID:9278503
A/Accession: F64776
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-397 <BAT>
A/Cross-references: GB:AE000152; GB:U00096; NID:G1786660; PIDN:AACT3565.1; PID:G1786668;
A/Experimental source: strain K-12, substrain MG1655
C/Genetics:
A/Gene: acra
A/Function:
A/Description: multidrug efflux pump
A/Note: active form contains proteins acra and acrc

C:Superfamily: lipoyl/biotin-binding homology
 C:Keywords: lipoprotein
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-397/Product: acriflavine resistance protein acra #status predicted <MAT>
 F:61-104,176-205/Domain: lipoyl/biotin-binding homology #status atypical <LPB>

Query Match 37.8%; Score 762.5; DB 2; Length 397;
 Best Local Similarity 42.2%; Pred. No. 2,8e-39;
 Matches 168; Conservative 76; Mismatches 129; Indels 25; Gaps 6;

Qy 11 AALAAALVALVSSCGKGDAAGGQAPAGREAPAVGVVTHPQTVALTYELPERLESL 70
 Db 11 AVLMISGLALATGCD-DKQAQGGQ-----QMAVGVVTVKTEPLDITTELPRTAY 63
 Qy 71 RTADVRAVGGIIGRLFOEGSVYRAGPLVQIDSTYEANLESARALATQAATLAKAD 130
 Db 64 RIAEVRPVSGLILKRNKESGDIAGVSLVQIDPATYQATYDSAKGLARAQAANATQ 123
 Qy 131 ADLARYPELVAAEAVSRQEDAAVTAKRSAEAGVAAQAIAKSAGININRSRTAPISGF 190
 Db 124 LTVNRQKLTGTQYISKQEDQALADAAQANAAVAATAAAYETARINLAYTKVTSPIGR 183
 Qy 191 IGGSVSEGTLLNAGDTTVLATIRQTNPMYVNVQSAEYWKLRQIAEGKLLAADGVIA 250
 Db 184 IGKSVTEGALVQNGAALATVQQLDPIYDVVTQSSNDPRLKQELANGTLKQENGAK 243
 Qy 251 VGIRFDGTVPYPEKGRLLFADPVNVESTGOITLRAAVNDONILMPGLYVRVLMDOVAVD 310
 Db 244 VSLTSDGIKPFQDGTLEFSDVTYVQTIGSTILRAIFPNPDHTLLPGMFVPAKLEGINP 303
 Qy 311 NAFVVPQQAATVRGAKD--TVMIVNAGGMEPREVTVAQOQGTNMTVITSLGKDGKRVVEG 368
 Db 304 NALIVPQGGVTRTPRGDAATLVVGADDKVETRPVIAAQIGDKMLVTDLGKAGDRVVISG 363
 Qy 369 ISIGITGAKKVT-----KEMASSENQAAPQSGVQ 400
 Db 364 L-----QKVRPQGVQVKAQEVITADNNQQA--SGAQ 391

RESULT 8

AG0561
 acriflavine resistance protein A precursor [imported] - Salmonella enterica subsp. enterica
 C:Species: Salmonella enterica subsp. enterica serovar Typhimurium
 A:Note: this species has also been called Salmonella typhi
 C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
 C:Accession: AG0561
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Garra, P.
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Nature 413, 848-852, 2001
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar
 A:Reference number: AB0502; MUID:21534947; PMID:11677608
 A:Accession: AG0561
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-397 <PAR>
 A:Cross-references: GB:AL513382; PIDN:CAD04961.1; PID:q16501748; GSPDB:GN00176
 C:Genetics:
 A:Gene: STY0520

Query Match 37.4%; Score 756; DB 2; Length 397;
 Best Local Similarity 41.3%; Pred. No. 7e-39;
 Matches 166; Conservative 78; Mismatches 129; Indels 30; Gaps 5;

Qy 11 AALAAALVALVSSCGKGDAAGGQAPAGREAPAVGVVTHPQTVALTYELPERLESL 70
 Db 11 AVLMISGLALATGCD-DKQAQGGQ-----QMEVGVVTVKTEPLDITTELPRTAY 63
 Qy 71 RTADVRAVGGIIGRLFOEGSVYRAGPLVQIDSTYEANLESARALATQAATLAKAD 130
 Db 64 RIAEVRPVSGLILKRNKESGDIAGVSLVQIDPATYQATYDSAKGLARAQAANAT 123

Qy 131 ADLARYPELVAAEAVSRQEDAAVTAKRSAEAGVAAQAIAKSAGININRSRTAPISGF 190
 Db 124 LTVNRQKLTGTQYISKQEDQALADAAQANAAVAATAAAYETARINLAYTKVTSPIGR 183
 Qy 191 IGGSVSEGTLLNAGDTTVLATIRQTNPMYVNVQSAEYWKLRQIAEGKLLAADGVIA 250
 Db 184 IGKSVTEGALVQNGAALATVQQLDPIYDVVTQSSNDPRLKQELANGTLKQENGAK 243
 Qy 251 VGIRFDGTVPYPEKGRLLFADPVNVESTGOITLRAAVNDONILMPGLYVRVLMDOVAVD 310
 Db 244 VSLTSDGIKPFQDGTLEFSDVTYVQTIGSTILRAIFPNPDHTLLPGMFVPAKLEGINP 303
 Qy 311 NAFVVPQQAATVRGAKD--TVMIVNAGGMEPREVTVAQOQGTNMTVITSLGKDGKRVVEG 368
 Db 304 NALIVPQGGVTRTPRGDAATLVVGADDKVETRPVIAAQIGDKMLVTDLGKAGDRVVISG 363
 Qy 369 ISIGITGAKKVT-----KEMASSENQAAPQSGVQ 397
 Db 364 L-----QKVRPQGVQVKAQEVITADNNQQA--SGAQ 391

RESULT 9

S39629
 multidrug-efflux transport protein A precursor - Pseudomonas aeruginosa
 N:Alternate names: multidrug resistance protein A
 C:Species: Pseudomonas aeruginosa
 C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 31-Dec-2000
 C:Accession: S39629; D83593
 R:Pooler, K.; Heinrichs, D.E.; Neshat, S.
 M01. Microbiol. 10, 529-544, 1993
 A:Title: Cloning and sequence analysis of an EnvC homologue in Pseudomonas aeruginosa
 A:Reference number: S39629; MUID:95058196; PMID:7968531
 A:Accession: S39629
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-383 <POO>
 A:Cross-references: GB:L11616; NID:9438852; PIDN:AAA74436.1; PID:9438853
 A:Experimental source: strain CD10
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Brice, S.; Young, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapidis, K.; Lim, J.; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
 A:Reference number: AB2950; MUID:20437337; PMID:10984043
 A:Accession: D83593
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-383 <STO>
 A:Cross-references: GB:AE004479; GB:AE004091; NID:9946272; PIDN:AA03814.1; GSPDB:GN00176
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: mexA; PA0425
 C:Function:
 A:Description: probably involved in secretion of the siderophore pyoverdine
 A:Note: inducible under conditions of iron limitation
 C:Superfamily: lipoyl/biotin-binding homology
 C:Keywords: lipoprotein
 F:1-23/Domain: signal sequence #status predicted <SIG>
 F:24-383/Product: multidrug-efflux transport protein A #status predicted <MAT>
 F:59-102,160-188/Domain: lipoyl/biotin-binding homology #status atypical <LPB>

Query Match 37.3%; Score 753; DB 2; Length 383;
 Best Local Similarity 42.1%; Pred. No. 1e-38;
 Matches 174; Conservative 68; Mismatches 125; Indels 46; Gaps 10;

Qy 8 AMRAALAAVALVSSCGKGDAAGGQAPAGREAPAV-----GVVTVHPQTVALTVE 62
 Db 6 AMRLVLPALVAIALSGCGKS-----EAPPAQTEPVGIVTLEAQTTLNTE 53
 Qy 63 LPRGLBELRTADVAVQVGGIIGRLFOEGSVYRAGPLVQIDSTYEANLESARALATA 122
 Db 54 LPRGTAIRIAEVRPQVNGIILKRLFKESGDVAKAQQLVQIDPATYQATYDSAKGLARAQAANLAST 113

QY 123 QATLAKADADLARKYPLVAEAASVROEYDAAVTAKESAAGVKAQAQAISAGINLNSR 182
 Db 114 Q-----EQARQKTLVADQAVSKQY-----ADANAAYLOSKAVEQARINLEKTV 159
 QY 183 ITAPISGFIGOSKVSSEGLTLAGDPTVATLRTNPMVNVNTOSASEVMKLRQIAGSKL 242
 Db 160 VLSISGRIGSASVTEGLVTVNGQANNAIVQQLDPIYDVTQSTALLRRRLASGQL 219
 QY 243 -LAADGVLAVGIKFPDDGTVPYPEKRLFPADPVNESTGOITLRAAVPNDQILMPLGLYVR 301
 Db 220 ERADBNNAKYSLKLEDSQYPLBGLRSEVSDEGTSVITRAVFRPNELLPGMFRVH 279
 QY 302 VLMDQAVDNAFVVPQQAATRGAKD--TVMVNAOGGMEPREVTVAOQOGTMMVITSGLK 359
 Db 280 AOLQEGVAKIILAPQOGVTRDLKGQATLAVVNAQNKVELFVIRKADRVIGDKMLVTGSLN 339
 QY 360 DGDVNVVVGISIA--GITGAKKVTPEKMASSENQAAPQSGVQVASEKTKASE 410
 Db 340 AGDKITIEGLQFVQPGVE-VKTVPAKNVASKAKADAP-----AKTDSK 382

RESULT 10

C65119

acriflavin resistance protein acrf precursor - Escherichia coli (strain K-12)
 N/Alternate names: envelope protein
 C/Species: Escherichia coli
 C/Date: 12-Sep-1997 #sequence revision 17-Sep-1997 #text_change 01-Mar-2002
 C/Accession: C65119; S18536; S18665
 R/Blatner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C.

A: Rose, D.J.; Mau, B.; Siao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617; PMID:9278503
 A:Accession: C65119
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-385 <BLAT>
 A:Cross-references: GB:AE000405; GB:U00096; NID:G1789659; PIDN:AACT6297.1; PID:G1789665;
 A:Experimental source: strain K-12, substrain MG1655
 R/Klein, J.R.; Henrich, B.; Plapp, R.
 Mol. Gen. Genet. 230, 230-240, 1991

A:Title: Molecular analysis and nucleotide sequence of the envC operon of Escherichia c
 A:Reference number: S18536; MUID:92079901; PMID:1720861
 A:Accession: S18536
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-190, 'AL', 193, 'SLM', 197, 'KRLNWLSSSILST', 213-312, 'T', 315-385 <KLE>
 A:Cross-references: EMBL:X57948
 R/Klein, J.R.; Henrich, B.; Plapp, R.
 Curr. Microbiol. 21, 341-347, 1990
 A:Title: Molecular cloning of the EnvC gene of Escherichia coli.
 A:Reference number: S18665
 A:Accession: S18665
 A:Status: preliminary; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-96 <KLE>
 A:Cross-references: EMBL:X57948; NID:9510827; PIDN:CAA41016.1; PID:9510829
 A:Experimental source: strain K-12
 C/Genetics:

A:Gene: acrf, envC
 C/Superfamily: lipoY/biotin-binding homology
 C/Keywords: cell division, lipoProtein
 F:1-23/Domain: signal sequence #status predicted <SIG>
 F:24-385/Product: acriflavin resistance protein acrf #status predicted <MAT>
 F:60-103, 175-204/Domain: lipoY/biotin-binding homology #status atypical <LPB>

Query Match 37.2%; Score 750.5; DB 2; Length 385;
 Best Local Similarity 42.1%; Pred. No. 1.4e-38;
 Matches 168; Conservative 74; Mismatches 122; Indels 35; Gaps 6;

QY 17 AVALVSSCGKGDAAGCGPAGREAPVVGVTVA--PQTVALTV--ELPGRLESIRTA 73
 Db 17 SAALVAGCNDKGEKAVHGEPO-----VTVAIVTAPLEVTLPGRTNAYRIA 65

QY 74 DVRAQVGIIQKRLFOEGSVYRAGQPLYOIDSSTYEANLESARAQLATAQATLAKADAL 133
 Db 66 EVRPQVSGIVLNNFTFGSDVQAGQSLYOIDPATYQANVDSAKELAKSEAAAIAHLTV 125
 QY 134 ARYKPLVAEASVROEYDAAVTAKESAAGVKAQAQAISAGINLNSRITTAISGTIGQ 193
 Db 126 KRYVPLVGTXYISQOEYDAIADARQADAIVAKAVESARINLAATYKTAPISGRIGK 185
 QY 194 SKVSEGLTLAGDPTVATLRTNPMVNVNTOSASEVMKLRQIAGSKLAAQGVAVGI 253
 Db 186 STVTEGLVTVNGQTTTELATVQQLDPIYDVTQSSNDPRLKQSVQEGNLHKNATSNVEL 245
 QY 254 KPDDGTVPYPEKRLFPADPVNESTGOITLRAAVPNDQILMPLGLYVRVLMDQAVDNAF 313
 Db 246 VMENGTVPYPLKGTLPQSDVTVDESTSITLRAVFPNQHLLPGRFARIDEGVQPDAL 305
 QY 314 VVPQQAATRGAKD--TVMVNAOGGMEPREVTVAOQOGTMMVITSGLKDGKRVVVGISI 371
 Db 306 LIPQOGVSRTPRGDAFVLIVNDSQVEVRPVVSAQIGDKMLISEGLKSGDQIVISGL-- 363
 QY 372 AGITGAKKVTPEKMASSENQAAPQSGVQVASEKTKASE 410
 Db 364 -----QKARPG-----QVATTDTIPADTASK 385

RESULT 11

A91146

hypotheical protein Ecs4137 [imported] - Escherichia coli (strain O157:H7, substrain R1M
 C/Species: Escherichia coli
 C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
 C/Accession: A91146
 R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;

Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genom
 A:Reference number: A99629; MUID:21156231; PMID:11258796
 A:Accession: A91146
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-385 <HAY>
 A:Cross-references: GB:BA000007; PIDN:BA037560.1; PID:G13363610; GSPDB:GN00154
 A:Experimental source: strain O157:H7, substrain R1MD 0509952
 C/Genetics:

A:Gene: Ecs4137
 Query Match 37.1%; Score 749.5; DB 2; Length 385;
 Best Local Similarity 42.1%; Pred. No. 1.7e-38;
 Matches 168; Conservative 74; Mismatches 122; Indels 35; Gaps 6;

QY 17 AVALVSSCGKGDAAGCGPAGREAPVVGVTVA--PQTVALTV--ELPGRLESIRTA 73
 Db 17 SAALVAGCNDKGEKAVHGEPO-----VTVAIVTAPLEVTLPGRTNAYRIA 65
 QY 74 DVRAQVGIIQKRLFOEGSVYRAGQPLYOIDSSTYEANLESARAQLATAQATLAKADAL 133
 Db 66 EVRPQVSGIVLNNFTFGSDVQAGQSLYOIDPATYQANVDSAGELAKSEAAAIAHLTV 125
 QY 134 ARYKPLVAEASVROEYDAAVTAKESAAGVKAQAQAISAGINLNSRITTAISGTIGQ 193
 Db 126 KRYVPLVGTXYISQOEYDAIADARQADAIVAKAVESARINLAATYKTAPISGRIGK 185
 QY 194 SKVSEGLTLAGDPTVATLRTNPMVNVNTOSASEVMKLRQIAGSKLAAQGVAVGI 253
 Db 186 STVTEGLVTVNGQTTTELATVQQLDPIYDVTQSSNDPRLKQSVQEGNLHKNATSNVEL 245
 QY 254 KPDDGTVPYPEKRLFPADPVNESTGOITLRAAVPNDQILMPLGLYVRVLMDQAVDNAF 313
 Db 246 VMENGTVPYPLKGTLPQSDVTVDESTSITLRAVFPNQHLLPGRFARIDEGVQPDAL 305
 QY 314 VVPQQAATRGAKD--TVMVNAOGGMEPREVTVAOQOGTMMVITSGLKDGKRVVVGISI 371
 Db 306 LIPQOGVSRTPRGDAFVLIVNDSQVEVRPVVSAQIGDKMLISEGLKSGDQIVISGL-- 363

Qy 372 AGITGAKKVTPEKEMASSENQAAPQSGVQTASEAKTASE 410
 Db 364 -----QKARPG-----QVKATDTDPADTAK 385

RESULT 12

hypochemical protein acce [imported] - Escherichia coli (strain O157:H7, substrain EDL93
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

C:Species: Escherichia coli
 C:Accession: E85991

R:Perma, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: AB5480; MUID:21074935; PMID:11206551

A:Accession: E85991

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-385 <STO>

A:Cross-references: GB:AE005174; NID:g12517895; PIDN:AA058393.1; GSPDB:GN00145; UMGF:246

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics: acce

Query Match 37.1%; Score 749.5; DB 2; Length 385;

Best Local Similarity 42.1%; Pred. No. 1.7e-38; Indels 35; Gaps 6;

Matches 168; Conservative 74; Mismatches 122; Indels 35; Gaps 6;

Qy 17 AVALVLSGCGKGDAAQGGPAGREAPVGVVTVH--PQTVALTV--ELPGRLESLRTA 73

Db 17 SAALAGCNDKGEKHAHGEPO-----VTVHIVKTAPLEVKTELPGRTNAYRIA 65

Qy 74 DVRAQVGGIIQKRLFOEGSVYRAGOPLYQIDSSYEAANLESARQAQATLAQATLAKADL 133

Db 66 EVRPQVSGIVANRNFTGSDVQAGSLQYQIDPATQAYNSAKGLAKSEAAALAHILTV 125

Qy 134 ARYPRVLAEEAVSRQEVAAATAKRSAGVKAQAALIKSAGINLNRSTAPISGFIQ 193

Db 126 KRIYPLVTKYISQOEYQALADARQADAAVIAAATVESAARINLAATKVTAPISGRIGK 185

Qy 194 SKVSEGLTLNAGDTTVALTIQTNPVYNTQSASEVWKLRRQJAEGLLAADYIAVGI 253

Db 186 STVREGALVTNGQTLTAVQQLDPIYDVTQSSNDPRLKQSVQGNLHKENATSNVEL 245

Qy 254 KFDQGVYPEKGRLLFADPVVNESTGQITLRAAVNDONILMPGLYFVLMDOVAVDNAF 313

Db 246 VMENQGTPLKGTLOFSDVTVESTGSLTLRAVFPNPQHTLLPGMFVRAARIDEGVQPDAI 305

Qy 314 VVPOOAVTRGAKD--TWIIVNAAQGMREPVTAQOQGTNWIIVTSGLDGPKVVEGISI 371

Db 306 LIPOQVSRTRGDAIVLIVDKSGVEVRPVVASAIDDKMLISEGLSGQGVIVSGV-- 363

Qy 372 AGITGAKKVTPEKEMASSENQAAPQSGVQTASEAKTASE 410

Db 364 -----QKARPG-----QVKATDTDPADTAK 385

RESULT 13

multidrug efflux protein [imported] - Yersinia pestis (strain CO92)
 C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001

C:Species: Yersinia pestis
 C:Accession: AD0380

R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
 11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,

Nature 413, 523-527, 2001

A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A:Reference number: AB0001; MUID:21470413; PMID:11586360

A:Accession: AD0380

A:Status: preliminary

A:Molecule type: DNA
 A:Residues: 1-395 <KIR>
 A:Cross-references: GB:AL590842; PIDN:CAC92367.1; PID:g15981070; GSPDB:GN00175
 C:Genetics: acce

Query Match 36.0%; Score 726.5; DB 2; Length 395;

Best Local Similarity 40.6%; Pred. No. 4.3e-37;

Matches 164; Conservative 75; Mismatches 142; Indels 23; Gaps 5;

Qy 11 AALAAAVLVLSGCGKGDAAQGGPAGREAPVGVVTVH--PQTVALTV--ELPGRLESL 70

Db 11 AAILVLSGSLVLCG--NDKQAAQAG-----AQQAPEVGVVTLKKEPINITTDLPGRSAF 64

Qy 71 RTADVAQVGGIIQKRLFOEGSVYRAGOPLYQIDSSYEAANLESARQAQATLAQATLAKAD 130

Db 65 RVAEVRPQVSGIILKRYNIEGSDVTAQSLYQIDPATQAYNSAKGLAKSEAAALAHILTV 124

Qy 131 ADLARYPLVLAEEAVSRQEVAAATAKRSAGVKAQAALIKSAGINLNRSTAPISGFIQ 190

Db 125 LTVNRKYPPLGNTNYSKQEDQALSDAQADATVLAALAKALESARINLAATQVNSPISGR 184

Qy 191 IGQKSEBGLTLNAGDTTVALTIQTNPVYNTQSASEVWKLRRQJAEGLLAADYIAVGI 250

Db 185 TGSAAVTEGALVTSQASAMTVQQLDPIYDVTQSSNDPRLKQSVQGNLHKENATSNVEL 244

Qy 251 VGIRFDQGVYPEKGRLLFADPVVNESTGQITLRAAVNDONILMPGLYFVLMDOVAVD 310

Db 245 VRLLENGVEYETETGLTFEGSVTVDETTSITLRAVFPNPQHTLLPGMFVRAARIDEGVQ 304

Qy 311 NAFVPOOAVTRGAKD--TWIIVNAAQGMREPVTAQOQGTNWIIVTSGLDGPKVVEG 368

Db 305 DALVPOQVSRTRGDAIVLIVDKSGVEVRPVVASAIDDKMLISEGLSGQGVIVSGV-- 364

Qy 369 ISIGITGAKKVTPEKEMASSENQAAPQSGVQTASEAKTASE 412

Db 365 L-----QKIRP-----GVEVKVQEVETDTPAFETAPADTAK 393

RESULT 14

RND multidrug efflux membrane permease [imported] - Agrobacterium tumefaciens (strain C5

C:Species: Agrobacterium tumefaciens

C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002

C:Accession: AE2950

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.

11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,

Science 294, 2317-2323, 2001

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; MUID:21608550; PMID:11743193

A:Accession: AE2950

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-404 <KIR>

A:Cross-references: GB:AE008689; PIDN:AL44019.1; PID:g17741579; GSPDB:GN00187

A:Experimental source: strain C58 (Dupont)

C:Genetics: acce

A:Map position: linear chromosome

Query Match 35.8%; Score 723; DB 2; Length 404;

Best Local Similarity 40.0%; Pred. No. 7.3e-37;

Matches 161; Conservative 83; Mismatches 143; Indels 16; Gaps 7;

Qy 11 AALAAAVLVLSGCGKGDAAQGGPAGREAPV--VGVVTVH--PQTVALTV--ELPGRLES 69

Db 9 AAVLLTGLIVL-----GCSDEQASABA--APPQAVKVAIVAPBELPITNELPGRIAP 59

Qy 70 LRTADVAQVGGIIQKRLFOEGSVYRAGOPLYQIDSSYEAANLESARQAQATLAQATLAKA 129

Db 60 TRLAEVRPRVSGIIIVRVEQSGSLVKEGDLVLRIDRAPQVRVDSAEGLTLRRQAOLQA 119

QY 130 DADLARYKPLVAEAVSRQEDYDAVTAARSAAGVKAQAQAIKSGINLNRSRITAPISG 189

Db 120 RQTADROOQLARNSVSGQGFEDNAIALAQADAEVAAVEAGVAKRLNLQYADVKAPISG 179

QY 190 FIGQKVSSEGLTLNAGDTTVALTRQTNPMYVNTQSAEVMKLRRQIAEGLLAADGVI 249

Db 180 VIGRARITGALVATSGSENLATIQQLDPIYADFPQAPADLRIRKALQDQGLMTGQNEA 239

QY 250 AVGKFPDGYTPPEKGRLLFPADPVNVESTGOITRAAVPNQNTLMPGLYRVLMDOYAV 309

Db 240 EVNLFPDGSRYKPYSGRLPFSBAVDETTGVTLRGEFPNPNGLLPQMYRVQIQGSIQ 299

QY 310 DNAFVVPQQAQVTR--GAKDTVMIVNAQGMPEPREVTVAQOQGTWIVTSGLKDGDKYVE 367

Db 300 KAAFAVVPQAQVORAGQASVLYVNAEDTVQKRVSVRSIGDKWVISEGLDDGDRVYAE 359

QY 368 GISLAGITGAKKVTPEKWEASSENQAAPQSGVQTASEAKTASE 410

Db 360 GFQKTA--PGA-KVKPEPMWSEPDVAAA--AGSEGAAPSETSSSE 398

RESULT 15

F98332
hypothetical protein AGR_L_3215 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C/Species: Agrobacterium tumefaciens
C/Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C/Accession: F98332
R/Goodner, B.; Hinkley, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tun
A/Reference number: A97359; MUID:21608551; PMID:11743194
A/Accession: F98332
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-427 <KTR>
A/Cross-references: GB:AE007870; PIDN:AAK90184.1; PID:g15160189; GSPDB:GN00170
A/Genetics:
A/Map position: linear chromosome

Query Match 35.8%; Score 723; DB 2; Length 427;
Best Local Similarity 40.0%; Pred. No. 7,8e-37;
Matches 161; Conservative 83; Mismatches 143; Indels 16; Gaps 7;

QY 11 AAALAAVALVLSGCGGDAAGGQAGREAPAP-VGVVTVHPQTVALTELPGRLS 69

Db 32 AAVALTGLIV-----GCSDEQASADA--APPGAIVKVAVKPEELPITNELPGRIAP 82

QY 70 LRTADVPAQVGGIIQKRLFOGGSYVRAGQPLYOIDSSTYEANLESARQAOLATAQATLAKA 129

Db 83 TRLAEVRPRVSGIIIVRVEQSGSLVKEGDLVLRIDRAPQVRVDSAEGLTLRRQAOLQA 142

QY 130 DADLARYKPLVAEAVSRQEDYDAVTAARSAAGVKAQAQAIKSGINLNRSRITAPISG 189

Db 143 RQTADROOQLARNSVSGQGFEDNAIALAQADAEVAAVEAGVAKRLNLQYADVKAPISG 202

QY 190 FIGQKVSSEGLTLNAGDTTVALTRQTNPMYVNTQSAEVMKLRRQIAEGLLAADGVI 249

Db 203 VIGRARITGALVATSGSENLATIQQLDPIYADFPQAPADLRIRKALQDQGLMTGQNEA 262

QY 250 AVGKFPDGYTPPEKGRLLFPADPVNVESTGOITRAAVPNQNTLMPGLYRVLMDOYAV 309

Db 263 EVNLFPDGSRYKPYSGRLPFSBAVDETTGVTLRGEFPNPNGLLPQMYRVQIQGSIQ 322

QY 310 DNAFVVPQQAQVTR--GAKDTVMIVNAQGMPEPREVTVAQOQGTWIVTSGLKDGDKYVE 367

Db 323 KAAFAVVPQAQVORAGQASVLYVNAEDTVQKRVSVRSIGDKWVISEGLDDGDRVYAE 382

QY 368 GISLAGITGAKKVTPEKWEASSENQAAPQSGVQTASEAKTASE 410

Db 383 GFQKTA--PGA-KVKPEPMWSEPDVAAA--AGSEGAAPSETSSSE 421

RESULT 16

RND multidrug efflux membrane fusion protein MexC precursor PA4599 [imported] - Pseudomo
C/Species: Pseudomonas aeruginosa
C/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C/Accession: AB3072
R/Stover, C.K.; Pham, X.O.; Errin, A.L.; Miroguchi, S.D.; Watterer, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim,
J.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A/Reference number: AB2350; MUID:20437337; PMID:10984043
A/Accession: AB3072
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-387 <STO>
A/Cross-references: GB:AE004873; GB:AE004091; NID:g9950839; PIDN:AG07987.1; GSPDB:GN001
C/Experimental source: strain PA01
A/Genetics:
A/Genes: mexC; PA4599

Query Match 34.6%; Score 699.5; DB 2; Length 387;
Best Local Similarity 41.5%; Pred. No. 1.9e-35;
Matches 162; Conservative 82; Mismatches 131; Indels 15; Gaps 9;

QY 10 AAALAAVALVLSGCGGDAAGGQAGREAPAPVGVVTVHPQTVALTELPGRLS 69

Db 9 RIGALMAVIA--LAGCG----PAERQEA--EMVLV--VEVLTVQAEPLALSELPGRIEP 60

QY 70 LRTADVPAQVGGIIQKRLFOGGSYVRAGQPLYOIDSSTYEANLESARQAOLATAQATLAKA 129

Db 61 VRLAEVRPRVSGIIIVRVEQSGSLVKEGDLVLRIDRAPQVRVDSAEGLTLRRQAOLQA 120

QY 130 DADLARYKPLVAEAVSRQEDYDAVTAARSAAGVKAQAQAIKSGINLNRSRITAPISG 189

Db 121 QARVRREPLVKIQAVSQQDFDTADLRSAEATRSQAQDLFTARLNLGVASTAPISG 180

QY 190 FIGQKVSSEGLTLNAGDTTVALTRQTNPMYVNTQSAEVMKLRRQIAEGLLAADGVI 249

Db 181 RIGALVTGALVQGGCATLMARIQQLDPIYADFPQAPADLRIRKALQDQGLMTGQNEA 239

QY 250 AVGKFPDGYTPPEKGRLLFPADPVNVESTGOITRAAVPNQNTLMPGLYRVLMDOYAV 309

Db 240 ALTRV--EGTPYERQGLQFADVAVDRTGQIALRGKFNPDGVLPLGMYRVRTTPGID 298

QY 310 DNAFVVPQQAQVTR--GAKDTVMIVNAQGMPEPREVTVAQOQGTWIVTSGLKDGDKYVE 367

Db 299 NQAILVPRQAVHSSDSQAVVMVVGADERAESVGVGVWQSGRMQITBGLPECDRIYV 358

QY 368 GISLAGITGAKKVTPEKWEASSENQAAPQSGVQTASEAKTASE 410

Db 359 G--LAAVQPGKIVPKPDGAQAQASPAQ 386

RESULT 17

E85757
probable efflux pump Z2509 [imported] - Escherichia coli (strain 0157:H7, substrain EDL9;
C/Species: Escherichia coli
C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C/Accession: E85757
R/Berna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,
H.; Lior, L.; Grobeck, E.O.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A/Reference number: AB5480; MUID:21074935; PMID:11206551
A/Accession: E85757
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-373 <STO>
A/Cross-references: GB:AE005174; NID:g12515494; PIDN:AA656521.1; GSPDB:GN00145; UMGp:Z250

A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z2509

Query Match 34.5%; Score 697.5; DB 2; Length 373;
Best Local Similarity 42.7%; Pred. No. 2.3e-35;
Matches 162; Conservative 66; Mismatches 134; Indels 17; Gaps 7;

Qy 9 MRAAALAAVALVSSCGKGDAAAGGCPAGREAPAPVGVTVHPQTALVELPGLE 68
Db 1 MKYIATSVVAMLLISGC-----DNTOSNNSSPSETE---VGVTTVSQQVSVSELTGRTS 53
Qy 69 SLRTADVAAOVGGIIQKRLFOEGSVYRAGQPLVYIDSSSTEYANLESAPALATQAATLAK 128
Db 54 AALSSEVAPQVGGIIQKRLFKEGDLVKAGQPLVYIDASSYQAAWNEAPALQQAQ-A-LVK 112
Qy 129 ADADLA-RYKPLVAAEAVSRQETDAAVTAKSAEAGVAAQAQAIKSAGINLRSTITAPI 187
Db 113 ADCOKAQRYARLVKENGVSQODADDAOSTCAQDKASVAAKKALETARINLDMVTITAPI 172
Qy 188 SGFTGQSVSEGTLLNAGDTTVALTIRQTNPMYVNTQSASEVMKLRRQIAEGKLLAADG 247
Db 173 SGRIGISSVTPGALVTASQDALTTRIGLDIMTVDRSSVDLRLRKQ---SLATNSD 228
Qy 248 VIAVGIFDQGVYPEKGRLLFADPVNESTGQITLRAAVPNDONILMPGLYVRLMDQV 307
Db 229 TMSVSLIEDGTYSSEKRLLEVAVDESGVTLRAIFPNPQQQLPGMFVRAVDEG 288
Qy 308 AVDNAFVVPQQAIVRGAKD--TWMIVNAAGMEPREVTVAAQQGNTNIVTSGLKXGDKRV 365
Db 289 VMEAIILAPQOGVTRDAKGNATLVNKNKVEORTLETGETYGDKMLVINGLSGDRLI 348
Qy 366 VEGISIGITGAKKVTPE 384
Db 349 VEG--SAKVTSGQTVKAVE 365

RESULT 18

G90861
probable efflux pump Ecs1863 [similarity] - Escherichia coli (strain O157:H7, substrain C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 02-Nov-2001
C:Accession: G90861
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G. gasawara, N.; Yasunaga, T.; Kuhara, S.; Shibata, T.; Hattori, M.; Shinagawa, H. DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc A:Reference number: A59629; PMID:21156231; PMID:11258796
A:Accession: G90861
A:Molecule type: DNA
A:Status: preliminary
A:Residues: 1373 <HAY>
A:Cross-references: GB:BA00007; PIDN:BAB35286.1; PID:g13361328; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 050952
C:Genetics:
A:Gene: Ecs1863

Query Match 34.5%; Score 697.5; DB 2; Length 373;
Best Local Similarity 42.7%; Pred. No. 2.3e-35;
Matches 162; Conservative 66; Mismatches 134; Indels 17; Gaps 7;

Qy 9 MRAAALAAVALVSSCGKGDAAAGGCPAGREAPAPVGVTVHPQTALVELPGLE 68
Db 1 MKYIATSVVAMLLISGC-----DNTOSNNSSPSETE---VGVTTVSQQVSVSELTGRTS 53
Qy 69 SLRTADVAAOVGGIIQKRLFOEGSVYRAGQPLVYIDSSSTEYANLESAPALATQAATLAK 128
Db 54 AALSSEVAPQVGGIIQKRLFKEGDLVKAGQPLVYIDASSYQAAWNEAPALQQAQ-A-LVK 112
Qy 129 ADADLA-RYKPLVAAEAVSRQETDAAVTAKSAEAGVAAQAQAIKSAGINLRSTITAPI 187
Db 113 ADCOKAQRYARLVKENGVSQODADDAOSTCAQDKASVAAKKALETARINLDMVTITAPI 172

Qy 188 SGFTGQSVSEGTLLNAGDTTVALTIRQTNPMYVNTQSASEVMKLRRQIAEGKLLAADG 247
Db 173 SGRIGISSVTPGALVTASQDALTTRIGLDIMTVDRSSVDLRLRKQ---SLATNSD 228
Qy 248 VIAVGIFDQGVYPEKGRLLFADPVNESTGQITLRAAVPNDONILMPGLYVRLMDQV 307
Db 229 TMSVSLIEDGTYSSEKRLLEVAVDESGVTLRAIFPNPQQQLPGMFVRAVDEG 288
Qy 308 AVDNAFVVPQQAIVRGAKD--TWMIVNAAGMEPREVTVAAQQGNTNIVTSGLKXGDKRV 365
Db 289 VMEAIILAPQOGVTRDAKGNATLVNKNKVEORTLETGETYGDKMLVINGLSGDRLI 348
Qy 366 VEGISIGITGAKKVTPE 384
Db 349 VEG--SAKVTSGQTVKAVE 365

RESULT 19

B82600
precursor of drug resistance protein XF2093 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: B82600
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; PMID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: B82600
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-408 <SIM>
A:Cross-references: GB:AE004025; GB:AE003849; NID:G9107217; PIDN:AF84892.1; GSPDB:GN001 A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A Britones, W.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrel, H as-Neto, E.; Docena, C.; El-Dorri, H.; Facincant, A.P.; Ferreira, A.J.S. submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigr chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A Rodrigues, V.; Rosa, A.J.; de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF2093

Query Match 33.2%; Score 670; DB 2; Length 408;
Best Local Similarity 36.8%; Pred. No. 1.2e-33;
Matches 146; Conservative 90; Mismatches 149; Indels 12; Gaps 5;

Qy 9 MRAAALAAVALVSSCGKGDAAAGGCPAGREAPAPVGVTVHPQTALVELPGLE 68
Db 17 LRVGLAVTVLALTACKS-----SEQP---QMPTEVSVVEAKPQTTPIQLDLVGRLS 67
Qy 69 SLRTADVAAOVGGIIQKRLFOEGSVYRAGQPLVYIDSSSTEYANLESAPALATQAATLAK 128
Db 68 AYRSADYARVAGLQRIYEGTEVEKQPLFQIDPAPFOATLLEAGRLAAAEATYKN 127
Qy 129 ADADLA-RYKPLVAAEAVSRQETDAAVTAKSAEAGVAAQAQAIKSAGINLRSTITAPI 188
Db 128 AKIVADRRRLSPQVYSRSDVDNAEAEKTPAASVQARPAAMONARINLNAVVTAPIS 187
Qy 189 GFIGOSKVSSEGTLLNAGDTTVALTIRQTNPMYVNTQSASEVMKLRRQIAEGKLLAADG 247
Db 188 GLAQOGKVTBEGALVSGSSSTLLITVDQIDPLVYVFNMSNDELMLRQAQHGSGVQLSSDN 247
Qy 248 VIAVGIFDQGVYPEKGRLLFADPVNESTGQITLRAAVPNDONILMPGLYVRLMDQV 307

Db 248 TSTIDVLLSDGSKVPHQGLIPSGCATVDPSTGTSLRAVLNPTDHLILPGAFTFKANLG 307
 Qy 308 AVDNAFVVPQQAATRGAKDT-VMIYNAOGMEPREVTAAOQGTNMTVTSGLKDGKVVV 366
 Db 308 QRRNTILLPGAGVQGDANASAVYLVDKQGVIRKNTVITGKNDQMLTAGLSNDRIV 367
 Qy 367 EGI-SIAGITGAKKVTTPKEMASSENQAAAPOSQVOTA 402
 Db 368 DGLQKVEKGAFAKAI PMKPDILISSAHAAKPTHTA 404

RESULT 20

S47733
 yHtu protein precursor - Escherichia coli (strain K-12)
 C:Species: Escherichia coli
 C:Date: 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 01-Mar-2002
 C:Accession: S47733; D65149
 R:Plunkett, G.
 submitted to the EMBL Data Library, March 1994
 A:Reference number: S47666
 A:Accession: S47733
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-385 <PLU>
 A:Cross-references: EMBL:U00039; NID:G466582; PIDN:AA18489.1; PID:G466650
 A:Experimental source: strain K-12, substrain MG1655
 R:Battner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
 A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617; PMID:9278503
 A:Accession: D65149
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-385 <BLAT>
 A:Cross-references: GB:AE000427; GB:U00096; NID:G1789919; PIDN:AACT6538.1; PID:G1789929;
 A:Experimental source: strain K-12, substrain MG1655
 C:Genetic:
 A:Gene: yHtu
 C:Superfamily: lipoyl/biotin-binding homology
 C:Keywords: lipoprotein
 F:1-20/Domain: signal sequence #status predicted <SIG>
 F:21-385/Product: yHtu protein #status predicted <MNT>
 F:57-100/172-201/Domain: lipoyl/biotin-binding homology #status atypical <LPB>

Query Match 32.5%; Score 656; DB 2; Length 385;
 Best Local Similarity 38.5%; Pred. No. 8.2e-33;
 Matches 143; Conservative 71; Mismatches 153; Indels 4; Gaps 2;

Qy 25 CGKGGDAAGGQGPAGREAPVAVGVTVHPQTVALTVELPGRLESLRTADVRAQVGGIIQ 84
 Db 14 CGAMLTACDDKSAENNTAAMTPEVGVVTLSPGSAVVLSELPGRTVPEVAELRPQVGII 73
 Qy 85 KRLFOEGSSYRAQGPLYQIDSSSTEYANLESARQALTAQATLAKADLARYKLVAAEA 144
 Db 74 KRNFIEDGKXNQGDSTLYQIDPAPLQAEINSAKSLAALSTASNAITFRNRQASLTKTNY 133
 Qy 145 VSRQEVDAATARKSAAGVKAQAATKSGININRSRITAPISGFIGOSKVSSEGTLLNA 204
 Db 134 VSRQDYATATQUNEAAANTVAKAAVEQATINIQYANVTSPITGVSGKSSVTGALVTA 193
 Qy 205 GDTTVALTIRQTNPMYVNTQASAEVWKLRLQALEGLAADGVIAVGIFPDGTVPYK 264
 Db 194 NQADSLVTQRLDPIYVDLTQVDFLRMEKSVASGQIKQVQSTPQVLNENKRSQT 253
 Qy 265 GRLLFADPVNVESTGQITTLRAAVNDONILMPGLYRVLMDOVAVDNAFVVPQQAATRG 324
 Db 254 GTLKFSDPYVDETTGVTLRAIFPNPGDILLPGMYVTALVDEGSRQVLLVPOEGVTHNA 313
 Qy 325 --KDTWIVNAOGMEPREVTVAOQGTNMTVTSGLKDGKVVVEGTS--IAGITGAKV 380
 Db 314 QGKATALLDKDQVVKLREIASKALGDQWVNTSGLOAGDVRIVSGLRIRPGIKARAI 373

Qy 381 TPKEWASSENQ 391
 Db 374 SSGENASTESK 384

RESULT 21

A91178
 probable membrane protein [imported] - Escherichia coli (strain O157:H7, substrain R1MD
 C:Species: Escherichia coli
 C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
 C:Accession: A91178
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
 gasawara, N.; Yasunaga, T.; Kuhnara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
 A:Reference number: A99629; MUID:21156231; PMID:11258796
 A:Accession: A91178
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-385 <HAY>
 A:Cross-references: GB:BA000007; PIDN:BA37816.1; PID:G13363867; GSPDB:GN00154
 A:Experimental source: strain O157:H7, substrain R1MD 050952
 C:Genetic:
 A:Gene: EC84393

Query Match 32.4%; Score 655; DB 2; Length 385;
 Best Local Similarity 38.5%; Pred. No. 9.4e-33;
 Matches 143; Conservative 71; Mismatches 153; Indels 4; Gaps 2;

Qy 25 CGKGGDAAGGQGPAGREAPVAVGVTVHPQTVALTVELPGRLESLRTADVRAQVGGIIQ 84
 Db 14 CGAMLTACDDKSAENNTAAMTPEVGVVTLSPGSAVVLSELPGRTVPEVAELRPQVGII 73
 Qy 85 KRLFOEGSSYRAQGPLYQIDSSSTEYANLESARQALTAQATLAKADLARYKLVAAEA 144
 Db 74 KRNFIEDGKXNQGDSTLYQIDPAPLQAEINSAKSLAALSTASNAITFRNRQASLTKTNY 133
 Qy 145 VSRQEVDAATARKSAAGVKAQAATKSGININRSRITAPISGFIGOSKVSSEGTLLNA 204
 Db 134 VSRQDYATATQUNEAAANTVAKAAVEQATINIQYANVTSPITGVSGKSSVTGALVTA 193
 Qy 205 GDTTVALTIRQTNPMYVNTQASAEVWKLRLQALEGLAADGVIAVGIFPDGTVPYK 264
 Db 194 NQADSLVTQRLDPIYVDLTQVDFLRMEKSVASGQIKQVQSTPQVLNENKRSQT 253
 Qy 265 GRLLFADPVNVESTGQITTLRAAVNDONILMPGLYRVLMDOVAVDNAFVVPQQAATRG 324
 Db 254 GTLKFSDPYVDETTGVTLRAIFPNPGDILLPGMYVTALVDEGSRQVLLVPOEGVTHNA 313
 Qy 325 --KDTWIVNAOGMEPREVTVAOQGTNMTVTSGLKDGKVVVEGTS--IAGITGAKV 380
 Db 314 QGKATALLDKDQVVKLREIASKALGDQWVNTSGLOAGDVRIVSGLRIRPGIKARAI 373
 Qy 381 TPKEWASSENQ 391
 Db 374 SSGENASTESK 384

RESULT 22

B86024
 probable membrane protein yHtu [imported] - Escherichia coli (strain O157:H7, substrain I
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: B86024
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,
 Miller, L.; Grobbeck, E.J.; Davis, N.W.; Llim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: B86024
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-385 <STO>

A:Cross-references: GB:AE005174; MID:g1518222; PIDN:AAG58654.1; GSPDB:GN00145; UWGP:249
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: ynfH

Query Match 32.4%; Score 655; DB 2; Length 385;
Best Local Similarity 38.5%; Pred. No. 9,4e-31;
Matches 143; Conservative 71; Mismatches 153; Indels 4; Gaps 2;

OY 25 CGKGGDAAGGQAGPAGREAPAPVGVVTPOTVALTVELPGRLESRTADVAQVGGIIQ 84
DB 14 CGAMLTACDDKSAENTAMTPREVGVITLSPGSVNVLSLPGRTVYEAIEIPQVGGII 73
OY 85 KRLFOGGSYVAGCPLYQIDSSTYEANLESAPQAQATLAKADADLARYKPLVAEA 144
DB 74 KRNFEIGKAVNGDGLYQIDAPLQALNSAKGLAKLSTASNRARITFNNQASLTKNY 133
OY 145 VSRQGYDAVTAKRSAEAGVKAQAQAIKASAGINLRSRTAPISGFIQSKVSGTLLNA 204
DB 134 VSRQGYDAVTAKRSAEAGVKAQAQAIKASAGINLRSRTAPISGFIQSKVSGTLLNA 193
OY 205 GDTTVALTIRQTNPMYVAVTOSASVMTLRQIAEGKLLADGVLANGIKPDGTVPREK 264
DB 194 NOADSLVTQRLDPIYVDTQSVQDFLRKKEVASGQIKQVQSGTTPVOLNLENGKRYQT 253
OY 265 GRLLFADPVVESTGQITLRAAVPNDONILMPGLYVRVLMQVAVDNAFVVPQCAVTRGA 324
DB 254 GTLKSDPTVDETTOSVTLRAFPNPNDLPLGMVTVLNVDEGSGQNVLVDPQEBVTINA 313
OY 325 --KDTVMIVNAGGMEPREVTVAAQOQGNWIVTSGLKDGKRVVVGIS--IAGITGAKV 380
DB 314 OGKATALLDKDDVYKLEIREASKAIGQWVVTSGLDGDRVYGLQIRIRGIRARAIS 373
OY 381 TPKEWASSENO 391
DB 374 SSQENASTESK 384

RESULT 23

AD0423
multidrug efflux protein [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C:Accession: AD0423
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titchell, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarrag, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Dougan, G.;
11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AD0423
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-386 <KOR>
A:Cross-references: GB:AL590842; PIDN:CAC92712.1; PID:g15981407; GSPDB:GN00175
C:Genetics:
A:Gene: YP03483

Query Match 31.1%; Score 628.5; DB 2; Length 386;
Best Local Similarity 37.5%; Pred. No. 3.9e-31;
Matches 146; Conservative 79; Mismatches 151; Indels 13; Gaps 5;

OY 9 MRAAALAAVAVLVSSCGKGGDAAGGQAGREAPAPVGVVTPOTVALTVELPGRLE 68
DB 1 MHSKILLACLVFTLVACDQ--SSSPSAPPSROE---VGAVTLTKQPVTLSSDLSGRTV 53
OY 69 SLRTADVAQVGGIIQKRLFOGGSYVAGCPLYQIDSSTYEANLESAPQAQATLAK 128
DB 54 AAMTSEVAPQVYDGIITKRLFTGSEVTVAGQVLYQIDPASVQAQVYATLAKALQNVVSKS 113
OY 129 ADADLARKPLVVAEAVSRQGYDAVTAKRSAEAGVKAQAQAIKASAGINLRSRTAPIS 188
DB 114 AKLAKORVYAAALAKENGVSQODADDAQTSYQOALNVAKETALLETARINLAVTVQFAPIS 173

OY 189 GFIGOSKVSSEGTLLNAGDTTVALTIRQTNPMYVAVTOSASVMTLRQIAEGKLLADGV 248
DB 174 GRISSTVTPGALVTAQNTTALATIRNLDPYVDTQSSAQQLALRRQOQAGNPTVANA- 232

OY 249 IAVGIRKDDGTVYPEKGRLLFADPVNVSTGQITLRAAVPNDONILMPGLYVRVLMQVA 308
DB 233 -PVOLTLEDGGSVVAHEGSLQTEVAVDATGAVTLRAKFPNBEHQLPGMFVRASVVRGV 291
OY 309 VDMAFVVPQCAVTRGAKD--TVMIVNAGGMEPREVTVAAQOQGNWIVTSGLKDGKRVV 366
DB 292 NNTIILAPQCGITHDAGNATLVVNOQOQVERREVERTERTIDSYMLISRLGLAAGDRLIV 351
OY 367 EGISIAGITGAKKVTPEKMWASSENOAAP 395
DB 352 EGTEKVSV--GDDVKFPEVSTTLPLVAEP 378

RESULT 24

E83393
RND multidrug efflux membrane fusion protein precursor PA2019 [imported] - Pseudomonas aeruginosa
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: E83393
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.V.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: E83393
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-396 <STO>
A:Cross-references: GB:AE004628; GB:AE004091; MID:g9948018; PIDN:AAG05407.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA2019

Query Match 31.1%; Score 628; DB 2; Length 396;
Best Local Similarity 38.7%; Pred. No. 4.3e-31;
Matches 155; Conservative 65; Mismatches 159; Indels 22; Gaps 5;

OY 14 LAAVALVLSGCGKGGDAAGGQAGREAPAPV--VGAVTVTPOTVALTVELPGRLESRT 72
DB 13 LAAVALVFLGCEBAAD-----AKTAEAPAVGVIVARPAPIGITSLSLPGRLAVRQ 65
OY 73 ADVRAQVGGIIQKRLFOGGSYVAGCPLYQIDSSTYEANLESAPQAQATLAKADAD 132
DB 66 AEVRARVAGIYTRLYEAGQVVRAGTVLFOIDPAPLKAALDISRGALARAASHAAADK 125
OY 133 LARYPVAAEAVSRQGYDAVTAKRSAEAGVKAQAQAIKASAGINLRSRTAPISGFI 192
DB 126 LKRYADILIKRAISEREYTERQTDARQALQIASAKALEQARRLGVAVTAPIDGAR 185
OY 193 QSKVSEGTLLNAGDTTVALTIRQTNPMYVAVTOSASVMTLRQIAEGKLL--AADGVAV 251
DB 186 RALVTGALVGBDSPTLPTREQIDPIYVNFQAPAGEVAAQRAIRREGQGVADKQIAV 245
OY 252 GIKEDDGTVYPEKGRLLFADPVNVSTGQITLRAAVPNDONILMPGLYVRVLMQVAVDN 311
DB 246 RLVLADSEYPLAGELFLPSDLAVPGTDTIAMRLFPNPREHLLPGGVYQVRLGRAVPQ 305
OY 312 AFVVPQCAVTRGACDTVM--IVNAGGMEPREVTVAAQOQGNWIVTSGLKDGKRVVVGIS 370
DB 306 AITVRDALIRTAQSAVAVKVPNGKGLVEDVEVRADTLQGRWMIISRLKGGEWIVENNA 365
OY 371 IAGITGAKKVTPEKMWASSENOAAPQSGVOTASEAKTASEA 411
DB 366 -----QHAAGSSVQAVVRQPASADAPSPLAASRA 394

RESULT 25


```

OY 182 IITPISFISQSKSECTILNA-GDTVLATIROTNMVNVV---QSAEWKLRQI 237
OY 172 KYIAPISITQEOBVNESSLITGDASSLSLISVQJLDPVYVNEFSTDTTEAEIAKLK-- 228
OY 238 AEGKLIADG-VIANGIKFDDGTVPYPERGRIILFADPVNESTGOITLRAAPNPOINILMP 296
Db 229 AERGATGSDADRLKIKILFSGSKAYVDHCGTIDFTSSSLDETGTGLGVAAVENNHRILP 288
OY 297 GLYVRVIMDQAVADNAFVVPQOAVTRGAK-DTWMIVNAQSGMEPREVTVAAQOGTNMIVT 355
Db 289 GQFVRAEILDIQVDAITVPRKALMQSQAQGFVVYVNNKDNVNEVREPVYTGARELLKNDWLIS 348
OY 356 SGLMDGDVVVVEGISIAGITGAKKVTPEMASSSENGAAPQSGVGTASEAK 406
Db 349 QGLNSGDVITVEGV-IKAVPG-RPVQPV--VQGVYDDKQAQAGKEQQAADK 395

```

RESULT 28

C:Probable drug-resistance protein [imported] - Sinorhizobium meliloti (strain 1021) magp
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: G95375
R:Barrett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bowe
; Kalmann, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti
A:Reference number: A95262; MUID:21396509; PMID:11481452
A:Accession: G95375
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-388 <KUD>
A:Cross-references: GB:AA006469; PTDN:AAK65565.1; PID:914524048; GSEDB:GN00165
A:Experimental source: strain 1021, megaplasmid pSymA
R:Galibert, F.; Fhan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pela, D.; Chailu, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalmann, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebaull, P.; Vandenbol, M.; Vortroler, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: Sma1664
A:Genome: plasmid

Query Match	23.9%	Score 482;	DB 2;	Length 388;
Best Local Similarity	33.5%;	Pred. No 3.2e-22;		
Matches 135;	Conservative 81;	Mismatches 163;	Indels 24;	Gaps 12;

QY	6	FK-AMRAALAAVA-LVLS	CGKGGDAAGGQ	PAGREAPVVGVT	VHPQVALIVEL	63					
Db	2	EKTI	FRSVDPLV	GSGLTCSAGD	-VAQTG-----	PVGVMITVQVENSPAHF	50				
QY	64	PGRL	ESTRTADVRAV	OGGII	IOKRLFOEGS	VYRAGOPLY	QIDSSRYE	NTLEBARQ	LATAQ	123	
Db	51	VGR	VEALNAVDIR	RAV	BEGFLERRL	PABQ	NEKGGDL	FTLER	ITYELALEB	QMTL	110
QY	124	ATLAKAD	ADLAR	RYKPLV	MAEAVS	ROEYDA	ATAKRS	EAAGYKAA	QAAIKS	AGIM	183
Db	111	TNFD	NAERQ	LQNR	NAL-SQ	RTVSQ	AVIEESH	APAD	DIARAS	VL	169
QY	184	TAP	ISGFIG	QSKV	EEGTL	INAGD	TTVAL	TIQ	NTNPM	VYNTQ	241
Db	170	KAP	IDGR	IRGRA	AVS	GLVSP-S	SEPL	RAVVQ	TDPI	RVF	227
QY	242	L	LAADG	VI	AVIK	REDG	TVPE	KEGR	LLFAD	VN	301
Db	228	DEL	AKG-Y	AL	TKRL	RSN	GEPY	QSG	KL	EPF	286
QY	302	V	LMQ	VA	VDNA	FVY	PQ	AV	TRG	AKDT-V	360

Db 28 / V I V E P E R E R E P V P V G S Y E D R E R F P L V V D G E S P A A V R I R A S V O G N M V Y E S L O G 346

Qy 361 G D K V N V E G --- I S I A G I T G A K K V T P K E W A S S E N Q A A P Q S V Q 400

Db 347 G E K L I V E G L Q R V S P G A V N E A G S V S G C D - A A I D T P A P A P R L S S Q 388

RESULT 29

PND multi-drug efflux membrane fusion protein MexE precursor PA2493 [imported] - *Pseudomonas aeruginosa*
 C:Species: *Pseudomonas aeruginosa*
 C:Date: 15-Sep-2000 #sequence
 C:Revision: 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: E83335
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Watterner, P.; Hickey, M.J.; Br
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapidis, K.; Lim,
 J.; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen
 A:Reference number: AB2950; MUID:20437337; PMID:10984043
 A:Accession: E83335
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-614 <STO-
 A:Cross-references: GB:AE004676; GB:AE004091; NID:G9948552; PIDN:AGC05881.1; GSPDB:GN00177
 A:Experimental source: strain PA01
 C:Genetics:
 C:Gene: mexE; PA2493

```

Query Match Similarity 22.8%; Score 460; DB 2; Length 414;
Best local similarity 33.2%; Pred. No. 7, 7e-21;
Matches 135; Conservative 74; Mismatches 176; Indels 22; Gaps 9;

QY 14 LAAVALVLLSSCGKGDAAGGOPAGREAPAVGVVTVHPQTALVTLPGLRLESRTA 73
DB 13 LALAVALVLSACGKAPETTG-----MAARKVSAVEIEQPLNEMDEFTGRLEAPESV 65
QY 74 DVRAVGIIQKRLFOEGSVYPAQGPLVQIDISTYANLESRAQQLATQNTLAKDADL 133
DB 66 ELRPVSVYIDRVAHGEGLVKKGDLLFQIDRPFEAEVKRLERAOQQAAADARSVNEA 125
QY 134 ARYKPLVAAEAVSRQEDAAVTAKRSAEKGVAQAIAKSAGINLRSRITAPISGFIG 193
DB 126 QRGELRSLSMNLSALADARTTAQAEAKAAVAATQQLDPAARLNSEFRTIAPIDGAVSR 185
QY 194 SKVSEGITLLNAGDTVTLATIROTNPYVNVVTSQASEVMK--LRQIASEKLLAAGVIA 250
DB 186 AEVTAGNLVNSGE-TLLTTLVSTDKYAYAFDADERFELKVEVLEARQ--AGRDTRSESPPV 242
QY 251 VGKRDGDTVPPEKRRLLFADPVNVVESTGQITLRAVPDPDQNTLMPGLYRVVLMDQYAVD 310
DB 243 LGLSEEDGN--PHLGRDPLDNOVNPRTITIGRAVPDAKSGFTGLYRLKLVGSKTY 300
QY 311 NAFVVPQQAV-TRGAKDTVMIVNAQAGMEPREVTAAQOQGTNMIYVSLGDKGDKRVVEGI 369
DB 301 AATLIDKDAVGTDLCKEFLVLVDGDNKTYRFTVEMPKLEGRLIVASGLSKGDRIVVNGL 360
QY 370 SIAGITGAKKVTPK--EWASSENQ--AAPQSGVQYQASEAKTASEE 412
DB 361 Q-RVRPGMOVDPQKREMASADTLATLARLRQSGVGDSEPPKVAASRD 405

RESULT 30
G96007
Probable acriflavin resistance protein [imported] - Sinorhizobium meliloti (strain 1021)
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #ext_change 30-Sep-2001
C:Accession: G96007
R:Refman, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernandez
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endo
A:Reference number: A95842; MUID: 21396508; PMID: 11481431
A:Accession: G96007
A:Status: preliminary

```


A:Molecule type: DNA
 A:Residues: 1-373 <KUR>
 A:Cross-References: GB:AL591985; PIDN:CA949727.1; PID:G15141214; GSPDB:GN00167
 A:Experimental source: strain 1021, megaplasmid pSymB
 R:Galibert, F.; Fianu, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler, P.; Chaim, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.; L.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure, hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
 A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.
 A:Reference number: AB6039; MUID:21368234; PMID:11474104.
 A:Contents: annotation
 C:Genetic: C:Gene: acb; SMD21497
 A:Genome: plasmid

Query Match 22.8%; Score 459.5; DB 2; Length 373;
 Best Local Similarity 30.8%; Pred. No. 7.2e-21;
 Matches 118; Conservative 75; Mismatches 157; Indels 33; Gaps 7;

9 MRAAALAAVALVSSCG-----KGGDAAGGPGAPAPVAVGVVTHPQTVALT 60
 1 MRSRVLAAL-----SCGLTTTIGVAGNAYAOQPPPVNVAIPAAM-----DKES 48
 61 VELPGRLESLTADVRAOVGGIIQKRLFOESGVYRAGOPLYOIDSSTEYANLESARAOA 120
 49 VDLGKVVAVQKVIDRAVSGFLEKVPEDGOKVSAGTVLVQVEGCAARALQEI DSGIA 108
 121 TQAATLAKADADLARYKPLVAEAVSROEYDAVTAKSAAEYKAAQAIAKSAGININR 180
 109 AAEGORDIAVLERBRAQRLIATNTVAQATLDTANAOVKKAADLRLKSGKONELNLSY 168
 181 SRTAPISGFIOGKVSSEGLTNAGDTVLATIRGTNPMYVNVQASAEVWKLARQIAEG 240
 169 TKILAPFGVGLTVDVYDALV-APDSGLVTLRLDITVEFPVATSLYSYERKEG 227
 241 KLLAADGVIAVGIFKDDGVYPEKGRLLFADPVNVESTGQITLRAAVNDONIMPLGYV 300
 228 EMSGAN---VSITLPNTDYPEKGTIDFVASTVSGQIDYTVTAEFNPGTLLDGLTV 284
 301 RVLMDQVADNAFVPOQAVTRGAKDT-VMIVNAQGMPEPVVAQOQGNMTWITSGLK 359
 285 RVLVEQSPQDVLAIVPOQAVORDQGAFAVWVDANSKVELRVDVSRSSGCAVAVAKGLK 344
 360 DGDGVVVGISIAIGTAKKVTTP 382
 345 EGENVITEGVG-----KVRP 359

Db

RESULT 31
 T30829.
 hypothetical protein mexE - Pseudomonas aeruginosa
 C:Species: Pseudomonas aeruginosa
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
 C:Accession: T30829
 R:Kohler, T.; Miches-Hamehour, M.; Henze, U.; Gotoh, N.; Curry, L.K.; Pechere, J.C.
 Mol. Microbiol. 23, 345-354, 1997
 A:Title: Characterization of MexE-MexF-OprN, a novel positively regulated antibiotic eff
 A:Reference number: Z20891; MUID:97197179; PMID:9044268
 A:Accession: T30829
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-414 <KOH>
 A:Cross-References: EMBL:X99514; NID:G1707643; PIDN:CAA67866.1; PID:G1707644
 C:Genetics: A:Note: mexE

Query Match 22.1%; Score 446; DB 2; Length 414;
 Best Local Similarity 32.7%; Pred. No. 5.5e-20;
 Matches 133; Conservative 73; Mismatches 179; Indels 22; Gaps 9;

14 LAAAVLVLSGCGKGAAGGPGAPAPVAVGVVTHPQTVALTVELPGRLESLRTA 73

Db

13 LALAVALVLSAGCAPETTG-----MAAPKVSVAEVIQPLNWEDEFTGRLEAPSV 65
 74 DVRAOVGGIIQKRLFOESGVYRAGOPLYOIDSSTEYANLESARAOAATLAKADADL 133
 66 ELRRVSGYIDRVAFEGHGVKKDDLFQIDPRFEAEVKELEAOQARAAQARSVNEA 125
 134 ARYKPLVAEAVSROEYDAVTAKSAAEYKAAQAIAKSAGININRSRTAPISGFIGQ 193
 126 QGRSLRASNIAISLADARTTAQAEKAAVAAVQAQDADARLNLSPFRITAPIDGRVSR 185
 194 SKVSEGLTNAGDTVLATIRGTNPMYVNVQASAEVWKLARQIAEGKILAADGVIA 250
 186 AEVTAQNLVNSGECSPK.Y-STDKVAAYFPDADERVFLKVELARQ-AGDTRSESPPV 242
 251 VGIKFDGTVYPEKGRLLFADPVNVESTGQITLRAAVNDONIMPLGYVAVMDQVAVD 310
 243 LGLSSEBEN--PHLGRDLFDNOWNPRTGTRGAHVPDNAGKEFTPGLYVRLKLVGSKTY 300
 311 NAFVPOQAV-TRGAKDTVMIVNAQGMPEPVVAQOQGNMTWITSGLKPDGVVREGI 369
 301 AATILIKOAVGTDLGKFFVLVDGDNKTVYRTVEMGPKLBELRIVRSGLSKGDRIVNGL 360
 370 SIAGITGAKKVTTPK-EMASSENOA--AAPQGVQTAASEAKTASEAE 412
 361 Q--RVPRGMQVDPQKVMASADTLATLARLQSGVSEPPKVAASKD 405

Db

RESULT 32
 AG2881
 HlyD family secretion protein (imported) - Agrobacterium tumefaciens (strain C58, Dupont
 C:Species: Agrobacterium tumefaciens
 C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
 C:Accession: AG2881
 R:Wood, D.W.; Seubald, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
 ; Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 ster, E.W.
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A:Reference number: AB2577; MUID:21608550; PMID:11743193
 A:Accession: AG2881
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-396 <KUR>
 A:Cross-References: GB:AE008688; PIDN:AL43469.1; PID:G17740974; GSPDB:GN00186
 A:Experimental source: strain C58 (Dupont)
 C:Genetics: A:Gene: mexE
 A:Map position: circular chromosome

Query Match 21.9%; Score 442.5; DB 2; Length 396;
 Best Local Similarity 30.8%; Pred. No. 8.4e-20;
 Matches 123; Conservative 81; Mismatches 173; Indels 23; Gaps 9;

10 RAALAAVALVLSGCG-----KGGDAAGGPGAPAPVAVGVVTHPQTVALT 61
 7 RRLATGGIGIAGLSVAAGALFDDLPTRSNATASTPA--ETPAIPVVAKESADVWRME 64
 62 ELPGRLESLTADVRAOVGGIIQKRLFOESGVYRAGOPLYOIDSSTEYANLESARAOAT 121
 65 EFGSRLEAVDRVQIRSVNAQIKVHREBALVKEGPTLTPIDAPVQAAVAGAGGVAS 124
 122 AQTALAKADADLARYKPLVAEAVSROEYDAVTAKSAAEYKAAQAIAKSAGININRS 181
 125 ABAKVSIAKTELDRGRSLSDNRITISGDLDOQOSSPADAAEAOIAAARAAALTTAOLDGYT 184
 182 RTAPISGFIOGKVSSEGLTNAGDT-VLATIRGTNPMYVNVQASAEVWKLARQI--A 238
 185 EITAPVSGRGRILEITGNVVAAGSTSPALTTLVSNPITASFNASGVAKALAEIPKT 244
 239 EGKLLAADGV-IAVGIFKDDGVYPEKGRLLFADPVNVESTGQITLRAAVNDONIMPLG 297

Db 245 DGAFLPALEQIIVEIGTLLSDEGR--PIKGTLLHIDNQVVSAGTIGVRAIFPNPDRLIPG 302
 Qy 298 LTVRLMQVAVNDNAFVVPQAV--TRGAKDTVMYINAGSGEPREVTVAQOGGTWIVTS 356
 Db 303 QFVRKMEPEKPEKRNIVISDRAIGTDQKRFVYVDAENKVSRYKIKGAPADGQRITDS 362
 Qy 357 GIKDGDVVVEGISIAGITGAKKVTPEKMASSENQAAPO 396
 Db 363 GLAAGDTIVVNGLO--RIRPGATIAPO----AEDKVAASQ 396

RESULT 33

hypochemical protein AGR_C_4505 [imported] - Agrobacterium tumefaciens (strain C58, Cere
 C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
 C:Accession: G97657
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;
 Science 294, 2323-2328, 2001
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
 A:Reference number: A97359; MUID:21608551; PMID:11743194
 A:Accession: G97657
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-396 <KIR>
 A:Cross-references: GB:AE007869; PIDN:AAK88216.1; PID:g15157668; GSPDB:GN00169
 C:Genetics:
 A:Gene: AGR_C_4505
 A:Map position: circular chromosome

Query Match 21.4%; Score 442.5; DB 2; Length 396;
 Best Local Similarity 30.8%; Pred. No. 8.4e-20;
 Matches 123; Conservative 81; Mismatches 173; Indels 23; Gaps 9;

Qy 10 RAAALAAVALVLSGCG-----KGGDAAGCGGAPAGREAPVAVVTVHPQTALTV 61
 Db 7 RRALTGAGIGLAMSVAAGALFFDLPTSRNATVASTPA--ETPAPVTVAKESRDVWME 64
 Qy 62 ELPGRLSRLTADVAQVGGIIOKRLFOGGSYVRAQGLYQIDSTYANIESARQIAT 121
 Db 65 EFSGRLKLVNDVQIRSRVAGQIKAVHFREGALVKEGDLFTIDPAPYQAAVAGEGVAVS 124
 Qy 122 AATLAKADADLARYKPIVAEAVSROEYDAVATKRSAGEGVKAAAIKSGINLRS 181
 Db 125 AEAKVSLKTELDKRRISDNKRTISQSDLPDQROSPADAEQLARARALTALTAQLDLYT 184
 Qy 182 RITAPISGFIGOSKVSSEGLTLNAGDTT-VLATIRQTNPMYVNTQASSEWKLRRQI--A 238
 Db 185 EILNPVSGVRGRIETAGNLVAAGSTSPALTTIVSVNIFASFNASEGVAKALAEIPKT 244
 Qy 239 EGKLIADGV-IANGIKRDDGTVPKERRLLPADPVNESGQITLRAVNDQILMPG 297
 Db 245 DGAFLPALEQIIVEIGTLLSDEGR--PIKGTLLHIDNQVVSAGTIGVRAIFPNPDRLIPG 302
 Qy 298 LTVRLMQVAVNDNAFVVPQAV--TRGAKDTVMYINAGSGEPREVTVAQOGGTWIVTS 356
 Db 303 QFVRKMEPEKPEKRNIVISDRAIGTDQKRFVYVDAENKVSRYKIKGAPADGQRITDS 362
 Qy 357 GIKDGDVVVEGISIAGITGAKKVTPEKMASSENQAAPO 396
 Db 363 GLAAGDTIVVNGLO--RIRPGATIAPO----AEDKVAASQ 396

RESULT 34

hypochemical protein b2243 [imported] - Escherichia coli (strain O157:H7, substrain
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: E85834
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grochbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,

Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: E85834
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-464 <STO>
 A:Cross-references: GB:AE005174; NID:g12516280; PIDN:AA657137.1; GSPDB:GN00145; UWGP:232
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: Z3243

Query Match 21.4%; Score 432; DB 2; Length 464;
 Best Local Similarity 31.9%; Pred. No. 4.5e-19;
 Matches 118; Conservative 63; Mismatches 169; Indels 20; Gaps 6;

Qy 34 GGGPAGREAPVAVVTVHPQTALTVLPGRLSRLTADVAQVGGIIOKRLFOGGSY 93
 Db 98 GGRGMSGSLAPVQAATVAVQAVPRYLTGITTANTVTVRSRVGQIMALHFQEGXQ 157
 Qy 94 VRAGPIYQIDSTYANIESARQIATDAQADLARYKPIVAEAVSROEYDAA 153
 Db 158 VKAGDLAEIDPSFKVALAQOGQLKDKKATLTNARDLARYQQLAKTNLVSQDEDAQ 217
 Qy 154 VTAKRSAGVKAQAQAAIKSAGINLNSRITAPISGFIGOSKVSSEGLTLNAGDTTVLATI 213
 Db 218 QALVSEGTGKADQEAQVSAQQLDMSRITAPVDGRVGLKQVDVNGQISSGDTTGIIVI 277
 Qy 214 RQTNPMYVNTQASSEWKLRRQIABEGLLAADGVAVGI-KFPDGTVPKERRLLPADP 272
 Db 278 TQTHIDLLFTLPSDIATVQAOAKGPLVEAMDRNRSKSEGT-----LLSLDN 330
 Qy 273 VVNSTGOITLRAVPNDQNLMPGLVY--RVLMDOVAVNDNAFVVPQAAVTRGAK-DTVM 329
 Db 331 QIDATTGKIKKAFPNQDDLPFNQVFNARMLVD--TEQNAVVIPTALOMGEGHFVW 388
 Qy 330 IVNAQSGMEPREVTVAQOGGTWIVTSGLKDGKRVVEGI-----SIAGITGAKKVT 382
 Db 389 VLNSGNKVSXKLTVTPGIQDSQKIVIRAGISAGDRVTVGIDRLTEGAKVEVEAQSATTP 448
 Qy 383 KEWASSENQ 392
 Db 449 EKAATSRREYA 458

RESULT 35

hypochemical protein b2074 - Escherichia coli (strain K-12)
 C:Species: Escherichia coli
 C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
 C:Accession: A64974
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
 .A.; Rose, D.J.; Mau, B.; Snao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617; PMID:9278503
 A:Accession: A64974
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-464 <BLAT>
 A:Cross-references: GB:AE000297; GB:U00096; NID:g1788382; PIDN:AACT5135.1; PID:g1788389;
 A:Experimental source: strain K-12, substrain MG1655

Query Match 21.4%; Score 432; DB 2; Length 464;
 Best Local Similarity 32.2%; Pred. No. 4.5e-19;
 Matches 119; Conservative 63; Mismatches 168; Indels 20; Gaps 6;

Qy 34 GGGPAGREAPVAVVTVHPQTALTVLPGRLSRLTADVAQVGGIIOKRLFOGGSY 93
 Db 98 GGRGMSGSLAPVQAATVAVQAVPRYLTGITTANTVTVRSRVGQIMALHFQEGXQ 157
 Qy 94 VRAGPIYQIDSTYANIESARQIATDAQADLARYKPIVAEAVSROEYDAA 153

Db 158 VKAGDLAEIDPSQFKVLAQAQGLADKATLANARBDLARYOOLAKTNLVSROELDAQ 217
Qy 154 VTAKRSAEAGVAAQAIAKSGAGINLNSRRTAPISGFIGOSKVSSEGLLNAGDTTVLATI 213
Db 218 QALVSETEGRTKADBASVASAQLQIDMSRITAPVGRVGLKQVDGNGISSGDTTGLVI 277
Qy 214 ROTNPMYVNVVQASAEVWKLRQIAEGKLLADGVIANGI-KFDDGTVPPEKGRLLFPDP 272
Db 278 TQTHPIDLFTLPESDIAITVVOAQAKGPLVEAWDRTNKSLSEGT-----LISLDN 330
Qy 273 VNNESTQITLRAAVPNDQNIIMPEGLYV--RYLMDQVAVDNAFVVPQOAVTRGAK-DTVM 329
Db 331 QIDATTTIKVKARPNNDODLFPNQFNARMLVD--TEQNAVVIPTALQMGNEGHFW 388
Qy 330 IVNAQGMPEPREVTVAAQOGGTNMTVTSGLKDGDKVVEGI-----SIAGITGAKKVT 382
Db 389 VLNSNKVSKHLVTPGIDQSKVIRAGISAGDRVTDGIRLTGAKAVEVEAQSATTP 448
Qy 383 KEWASSENOA 392
Db 449 EEKATSREYA 458

RESULT 36

AC0771

Probable efflux system protein STY2339 [imported] - Salmonella enterica subsp. enterica
C/Species: Salmonella enterica subsp. enterica serovar Typh
A/Note: this species has also been called Salmonella typhi
C/Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C/Accession: AC0771
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher,
Th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Mout, S.; O'Gaora, P.
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AC0771
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-413 <PRA>
A:Cross-references: GB:AL513382; PIDN:CAD02489.1; PID:G16503353; GSPDB:GN00176
A:Genetics:
A:Gene: STY2339

Query Match 21.3%; Score 431; DB 2; Length 413;
Best Local Similarity 31.9%; Pred. No. 4.5e-19;
Matches 123; Conservative 64; Mismatches 170; Indels 28; Gaps 8;

Qy 32 AAGGAPAGREAPAVGVVTHPQVATLVLPGRLESRTADVRAQVGIIQKRLFOEG 91
Db 46 AAGRRGRMGROGFLAPVQAATATTQAVPRYLSGLGTVAANTVTVRSRVDDGLALAHQEG 105
Qy 92 SYRAGGPIYQIDSSSTEANLESRAQLATQAATLADADLARYKPLVAAEAVSROEYD 151
Db 106 QOVNAGDLAQIDPSQFKVLAQAQGLADKATLANARBDLARYOOLAKTNLVSROELD 165
Qy 152 AAVTAKSAEAGVAAQAIAKSGAGINLNSRRTAPISGFIGOSKVSSEGLLNAGDTTVLA 211
Db 166 AQAALVNETGRTKADBASVASAQLQIDMSRITAPVGRVGLKQVDGNGISSGDTTGLVI 225
Qy 212 TIRQTNPMYVNVVQASAEVWKLRQIAEGKLLADGVIANGIKEDGTVP-EGKRLFLFA 270
Db 226 VITQTHPIDLFTLPESDIAITVVOAQAKGPLVEA-----WDRTNKSLSEGLVLSTL 278
Qy 271 DPVNVSTQITLRAAVPNDQNIIMPEGLYV--RYLMDQVAVDNAFVVPQOAVTRGAK-DT 327
Db 279 DNOGIDPTTGTIKVKARPNNDODLFPNQFNARMLVD--TEQNAVVIPTALQMGNEGHFW 388
Qy 328 VMIYNAQGMPEPREVTVAAQOGGTNMTVTSGLKDGDKVVEGISAGITGAKKVTPEKMAS 387
Db 337 VWVLNENNVSKKRVKIGIDNNRVVISAGLSAGDRVVDGID--RLTEGAKV----- 387

Qy 388 SENQAAAPQSGVQTAPEAKTASEAE 412
Db 388 ---EVEPQT---TMADEKSPSRHE 406

RESULT 37

B90989

Probable membrane protein [imported] - Escherichia coli (strain O157:H7, substrain R1MD
C/Species: Escherichia coli
C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C/Accession: B90989
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kunihara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: B90989
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-464 <HAV>
A:Cross-references: GB:BA000007; PIDN:BA836305.1; PID:G13362351; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain R1MD 0509952
A:Genetics:
A:Gene: Ec82882

Query Match 21.3%; Score 431; DB 2; Length 464;
Best Local Similarity 31.9%; Pred. No. 5.2e-19;
Matches 118; Conservative 63; Mismatches 169; Indels 20; Gaps 6;

Qy 34 GCGAPAGEAPAVGVVTHPQVATLVLPGRLESRTADVRAQVGIIQKRLFOEGSY 93
Db 98 GGRGRMGSGFLAPVQAATAVEQAVPRYLTGITTANTVTVRSRVDDGLALAHQEGQ 157
Qy 94 VRAQGPPIYQIDSSSTEANLESRAQLATQAATLADADLARYKPLVAAEAVSROEYDAA 153
Db 158 VKAGDLAEIDPSQFKVLAQAQGLADKATLANARBDLARYOOLAKTNLVSROELDAQ 217
Qy 154 VTAKRSAEAGVAAQAIAKSGAGINLNSRRTAPISGFIGOSKVSSEGLLNAGDTTVLATI 213
Db 218 QALVSETEGRTKADBASVASAQLQIDMSRITAPVGRVGLKQVDGNGISSGDTTGLVI 277
Qy 214 ROTNPMYVNVVQASAEVWKLRQIAEGKLLADGVIANGI-KFDDGTVPPEKGRLLFPDP 272
Db 278 TQTHPIDLFTLPESDIAITVVOAQAKGPLVEAWDRTNKSLSEGT-----LISLDN 330
Qy 273 VNNESTQITLRAAVPNDQNIIMPEGLYV--RYLMDQVAVDNAFVVPQOAVTRGAK-DTVM 329
Db 331 QIDATTTIKVKARPNNDODLFPNQFNARMLVD--TEQNAVVIPTALQMGNEGHFW 388
Qy 330 IVNAQGMPEPREVTVAAQOGGTNMTVTSGLKDGDKVVEGI-----SIAGITGAKKVT 382
Db 389 VLNSNKVSKHLVTPGIDQSKVIRAGISAGDRVTDGIRLTGAKAVEVEAQSATTP 448
Qy 383 KEWASSENOA 392
Db 449 EEKATSREYA 458

RESULT 38

A83330

Probable RND efflux membrane fusion protein precursor PA2528 [imported] - Pseudomonas aer
C/Species: Pseudomonas aeruginosa
C/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C/Accession: A83330
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Bro
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
J.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: AB2950; MUID:20437337; PMID:10984043
A:Accession: A83330
A:Status: preliminary
A:Molecule type: DNA

• Tue Sep 9 08:55:33 2003

us-09-889-756a-2.rpr

Page 17

Search completed: September 8, 2003, 14:02:44
Job time : 24 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2003, 14:02:17; Search time 26 Seconds
(without alignments)
2175.735 Million cell updates/sec

Title: US-09-889-756A-2

Perfect score: 2019

Sequence: 1 MAFYAFKMAAALAAVAL.....AAPSQVGTASEAKTASEAE 412

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 513375 seqs, 137303645 residues

Total number of hits satisfying chosen parameters: 513375

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*

2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*

3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*

4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*

5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*

6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*

7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*

8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*

9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*

10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*

11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*

12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*

13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*

14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*

15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*

16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*

17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*

18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	432	21.4	464	9	US-09-912-020-281
2	370.5	18.4	202	11	US-09-975-719-149
3	175	8.7	659	10	US-09-738-626-6815
4	169	8.4	299	9	US-09-815-242-10168
5	169	8.4	309	9	US-09-815-242-13850
6	161.5	8.0	285	9	US-09-815-242-11913
7	156.5	7.8	390	9	US-09-815-242-11146
8	151.5	7.5	399	12	US-09-769-744A-94
9	138.5	6.9	355	9	US-09-741-669-435
10	127.5	6.3	329	11	US-09-988-067B-52
11	123.5	6.1	329	8	US-08-945-038-4
12	123.5	6.1	407	9	US-09-784-208-4
13	122.5	6.1	407	14	US-10-078-107-2
14	122.5	6.1	407	14	US-10-077-751-2
15	122.5	6.1	407	15	US-10-315-023-4

16	122.5	6.1	407	15	US-10-315-023-10	Sequence 10, Appl
17	120.5	6.0	1180	12	US-10-193-764-61	Sequence 61, Appl
18	120.5	6.0	1188	12	US-10-193-764-59	Sequence 59, Appl
19	119.5	5.9	696	15	US-10-156-761-9338	Sequence 9338, Ap
20	118.5	5.9	407	15	US-10-315-023-11	Sequence 11, Appl
21	118	5.8	382	15	US-10-156-761-12596	Sequence 12596, A
22	118	5.8	1741	11	US-09-971-536-68	Sequence 68, Appl
23	116	5.7	607	11	US-09-308-207-59	Sequence 69, Appl
24	115	5.7	1208	15	US-10-156-761-13234	Sequence 13234, A
25	114.5	5.7	1208	15	US-10-156-761-13251	Sequence 13251, A
26	114	5.6	837	9	US-09-815-242-5883	Sequence 5883, Ap
27	114	5.6	875	9	US-09-815-242-13080	Sequence 13080, A
28	114	5.6	2434	9	US-09-815-242-5835	Sequence 5835, Ap
29	114	5.6	6281	9	US-09-815-242-12996	Sequence 12996, A
30	112.5	5.6	272	15	US-10-156-761-12370	Sequence 12370, A
31	112.5	5.6	405	9	US-09-815-242-10096	Sequence 10096, A
32	112.5	5.6	405	14	US-10-078-107-6	Sequence 6, Appl
33	112.5	5.6	405	14	US-10-077-751-6	Sequence 6, Appl
34	112	5.5	542	9	US-09-741-669-414	Sequence 414, App
35	112	5.5	578	9	US-09-159-469-50	Sequence 50, Appl
36	112	5.5	578	9	US-09-798-042-585	Sequence 585, Ap
37	111	5.5	1029	9	US-09-815-242-13083	Sequence 13083, A
38	111	5.5	1048	9	US-09-815-242-13083	Sequence 13083, A
39	111	5.5	1525	9	US-09-782-714-1	Sequence 1, Appl
40	111	5.5	7746	15	US-10-156-761-7965	Sequence 7965, Ap
41	109	5.4	542	11	US-09-308-207-16	Sequence 16, Appl
42	109	5.4	596	12	US-10-238-075-416	Sequence 416, Appl
43	108.5	5.4	338	10	US-09-881-752A-222	Sequence 222, App
44	108.5	5.4	482	12	US-10-017-161-1994	Sequence 1994, Ap
45	108.5	5.4	534	15	US-10-156-761-14197	Sequence 14197, A

ALIGNMENTS

RESULT 1

US-09-912-020-281

Sequence 281, Application US/09912020

Patent No. US2002004592A1

GENERAL INFORMATION:

APPLICANT: Zyskind, Judith

APPLICANT: Ohlsen, Kari L.

APPLICANT: Tirawick, John

APPLICANT: Forsyth, R. Allyn

APPLICANT: Froelich, Jamie M.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN

FILE REFERENCE: ELITRA 001DVI

CURRENT APPLICATION NUMBER: US/09/912,020

PRIOR FILING DATE: 2001-07-23

PRIOR APPLICATION NUMBER: 09/492,709

PRIOR FILING DATE: 2000-01-27

PRIOR APPLICATION NUMBER: 60/117,405

NUMBER OF SEQ ID NOS: 485

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 281

LENGTH: 464

TYPE: PRT

ORGANISM: E. Coli

US-09-912-020-281

Query Match 21.4%; Score 432; DB 9; Length 464;

Best Local Similarity 33.2%; Pred. No. 2.1e-30;

Matches 119; Conservative 63; Mismatches 168; Indels 20; Gaps 6;

QY 34 GGGPAGEAPAPVGVTVHPQYALTVELPGRLESRTADYRQVGGIIOKLFOEGSY 93

DB 98 GGRGRMSGPIAPVQATAVEQAVRYLTGIGITTAANTYVRSRVGQIIAIFQSGQQ 157

QY 94 VRAQGPVQIDSSYVEANLESARQALATQAATLAAADADLARYKPLVAEAVSROEYDAA 153
DB 158 VAAQGLLAIBDPSQFKVLAQAQOLADKATLANARRDLARYQQLAKTNLSROELDAQ 217
QY 154 VTAKSAAGVYKAAQAIKSNININRSRITAPISGFIGOSVSGTILNAGDTVLATI 213
DB 218 QALVSETGTTIADASVASAQLQIDMSRTAPVGRVGLKQVDVGNQISSGDTTGIIVI 277
QY 214 ROTNMYVNVVQOSASEVWKLRRQIEGKLLADGVIAVGI -KFDGDTVPKGRLLFADP 272
DB 278 TQTHIDLVTLPESDIAITVQAQAKPLVVEAMDRTNSKLSGCT-----LISLDN 330
QY 273 VNNESTGQITLRAAVPNDONIMPLVY--RVLMQVAVDNAFVVPQAVTRGAK-DTVM 329
DB 331 QIDATGTIKVARENNODDALFPNQFVNAKVLVD--TEQNAVVIPTAALQWNEGHFVW 388
QY 330 IVNAGGMEPREVYTAQOQGTWVYTSGLKQGDKVVVEGI-----SIAGITGAKVTP 382
DB 389 VLNSENKVSXKHLVTGIDQSQKRVIRAGISAGDRVVTGIDRLTEGAKVEVEAQSATTP 448
QY 383 KEMASSENQA 392
DB 449 EKATSRREYA 458

RESULT 2

US-09-975-719-149
; Sequence 149, Application US/09975719
; Publication No. US20030022349A1
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Rahme, Laurence G.
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; FILE REFERENCE: 00786/361003
; CURRENT FILING DATE: 2001-10-10
; PRIOR FILING DATE: 2001-10-10
; PRIOR FILING DATE: 1998-11-25
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 149
; LENGTH: 202
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-975-719-149

Query Match 18.4%; Score 370.5; DB 11; Length 202;
Best Local Similarity 37.9%; Pred. No. 2.1e-25;
Matches 80; Conservative 43; Mismatches 73; Indels 15; Gaps 5;
QY 205 GDTTVALITROTNPYVNVVQOSASEVWKLRRQIAEGKL-LAADGVIAVGIKFDGTVYPE 263
DB 1 GQAANAMTVOQLDIDYVDVTPSTALLMRRELASGQLERAGDMAAKSLKLEDSQYPL 60
QY 264 KGRLLFADPVNNESTGQITLRAAVPNDONIMPLGVYVVMQVAVDNAFVVPQAVTRG 323
DB 61 EGRLEFSVSVDREGSVTIRAVFPNNELLPGFVHAQLOEGVKQKATILAPQOGVTRD 120
QY 324 AKD--TVMAIVNAGGMEPREVYTAQOQGTNMIIVTSGLKQGDKVVVEGISIA--GITGAKK 379
DB 121 LKGAATLAVNAQNKVELRVIKADRVIGDKVLVTEGLNAGDKIITEGLOFVQGVVE-VKT 179
QY 380 VTPKEMASSENQAAAPQSGVQTASEAKTASE 410
DB 180 VPAKVVASQAKADAP-----AKTDSK 201

RESULT 3
US-09-738-626-6815
; Sequence 6815, Application US/09738626

; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OKAZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT FILING DATE: US/09/738, 626
; CURRENT FILING DATE: 2000-12-18
; PRIOR FILING DATE: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR FILING DATE: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR FILING DATE: JP 00/280988
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6815
; LENGTH: 659
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6815

Query Match 8.7%; Score 175; DB 10; Length 659;
Best Local Similarity 22.1%; Pred. No. 4.6e-07;
Matches 102; Conservative 81; Mismatches 194; Indels 84; Gaps 16;

QY 5 AFKMRRAALAAVALVSSCGKGDAAQCGPAGRAPAPVGVVTHPTVALTYBL- 63
DB 223 ALKARQADAAAEIERRADFGILNN-----DRSNLDVIGLDERESLASAESELA 274
QY 64 ---PGRLESRTADVAVQGGIIOKRLFOGSVYVRAGQPLVQIDSSYVEANLESARQA 119
DB 275 QARAQGLEAVAAAE--AKVAGLEQSIASKTSTPSPDQTYLQ---STRLEAEKRVYA 328
QY 120 ATAQA-----TLAKADADIARYKPLVAEAVSROEYDAV---TAKRSAAGVKA 166
DB 329 STTBALRIABRYIDSIGKVDSELAQAQRAV-AEASHAQ-DALGLETQLOSTQHOLEA 386
QY 167 AQAAIKSA-----GINNRKRTAPISGFI-----GQSKVSEGTILN 203
DB 387 QSSAIDAAALGLASVDNEAATRTSOLRMDINNTVRSYSGIVSVQAQGPAAAGALLS 446
QY 204 AGDPT---VLATITROTNPYVNV-----TQSASEVWKLRRQIAEGKLLAADGVIAVGIK 254
DB 447 VADSEIKITDANVEAEISNVTIGSRVTFPTTPTGTKEPGRVSKVSPILAA----- 497
QY 255 FDDGTVPEKGRLLFADPVNNESTGQITLRAAVPNDONIMPLGVYVVMQVAVDNAFV 314
DB 498 ---AASAPATGEGAAAGATTNTDVTPEIISVIGDEGLNLGGSARVRIYHETAPVLT 554
QY 315 VPQAVTRG--AKDTMIIVNAGGMEPREVYTAQOQGTNMIIVT-SGLKQGDKVVVEGISI 371
DB 555 VPLEAVYKNDGKXAVIILISDNNKVEVEVEKVTAESDPDFDAVSGAGISEBARVLTQGNV 614
QY 372 AGITGAKKVTVPKEMASSENQAAAPQSGVQTASEAKTASEA 412
DB 615 RGLIG---TVKLHADTVBQAAAFSPADPPDPAVAVSAK 652

RESULT 4
US-09-815-242-10168
; Sequence 10168, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: HaeiDeck, Robert


```

; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10168
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10168

```

```

Query Match      8.4%; Score 169; DB 9; Length 299;
Best Local Similarity 25.4%; Pred. No. 5e-07;
Matches 71; Conservative 43; Mismatches 105; Indels 60; Gaps 9;

```

```

QY 52 VHPQVTLVTVLPRLESRTADYRAQVGGIIOKRLFGESYVAGQPLYQIDSTYEAN 111
DB 54 IRAQVSTITPOVSGRIYEALNKD-----NOLVAGDILLTIDKTPFOIA 97
QY 112 LESARQATLTAQATLADADLARYKPLVAAEAVSROEYDAVAVTAKESAENGVAQAQAI 171
DB 98 ELNQAQALAKAQSDLAKANNEANRRHL-SQNPISAEELDTANLVAMQASVDAAQATL 156
QY 172 KSAQINLNRSKITPISGFIGQSVSESTLINAQDTVLATIRQTNPMYVNTQSASEVM 231
DB 157 KQAGWOLAQTEIRAPVSGWVNTLTTRIGDYADTG-----KPLFALVDSHSFYVI 205
QY 232 -----KLRRQIAGK-----LLAADG-----VIAVGIKPDGTVYPEKGRLLFPADPVN 275
DB 206 GYFEETKL-RHIRGAPAQITLYSDNKTLOGHVSIGALYDQVESDSSILT---PDVK 260
QY 276 ESTGQITLRAAVPNDONILMPGLYVRVLMDOVAVDNAFV 314
DB 261 PNVFVWRLAQRP-----VFALDKVPDGVTLV 288

```

```

RESULT 5
US-09-815-242-13850
; Sequence 13850, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in

```

```

; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13850
; LENGTH: 309
; TYPE: PRT
; ORGANISM: Salmonella typhi
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(309)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-13850

```

```

Query Match      8.4%; Score 169; DB 9; Length 309;
Best Local Similarity 24.5%; Pred. No. 5.2e-07;
Matches 63; Conservative 45; Mismatches 103; Indels 46; Gaps 7;

```

```

QY 75 VRAQVGGIIOKRLFGESYVAGQPLYQIDSTYEANLESARQATLTAQATLAA--DAD 132
DB 70 VTPQVSGSITQNLNKDQFNAGDVLFVIDKTPHIALNQAQALAKAQSDLAKANNEAD 129
QY 133 LARYKPLVAAEAVSROEYDAVAVTAKESAENGVAQAQAIKSAQINLNRSRTAPISFIG 192
DB 130 RRRH---LSRNYISAEELDSANLVAMQASVDVALLTLOQWOLSQTEVKAFVSGWVT 186
QY 193 QSKVSESTLINAQDTVLATIRQTNPMYVNTQSASEVM-----KLRR-----QI 237
DB 187 NLSTRIDYASTG-----KPLFALVDSHSFYVMGYFEETKLRIHIREGPALITL 235
QY 238 AEGKTLAADGVIAVGIFDGTVPYPEKGRLLFPADPVNNESTGQITLRAAVPNDONILMPG 297
DB 236 YSGNVKLOGHVGSGIRAIYDQVESDGLV---PDIKPVVWRLAQRP----- 282
QY 298 LVYRVLMDOVAVDNAFV 314
DB 283 --VRIEFDALPDITLV 297

```

```

RESULT 6
US-09-815-242-11913
; Sequence 11913, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21

```

PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11913
LENGTH: 285
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-815-242-11913

Query Match 8.0%; Score 161.5; DB 9; Length 285;
Best Local Similarity 26.0%; Pred. No. 2.2e-06;
Matches 69; Conservative 40; Mismatches 101; Indels 55; Gaps 10;

QY 71 RTADVRA-----OVGGIIQKRLFOEGSYVAGQPLVQIDSTYEANLESARQLATAQ 123
DB 35 RDARFRAVYVVPVPSGVWTDLEVKDQVYKGVLRIDDERQANLEQARAFAETRH 94
QY 124 ATLAKADADLARYPEL-----VAAEAVSRQEDYDAVATKRSAGYKAAQAAIKSAGINLN 179
DB 95 QOYLLRQNEAARSRSLGIGALISAEDEKENVQINAAI-----ASEYQEAALQVKAIELNLK 149
QY 180 RSRITAPISGRTGQSKVSEGTILNAGDTTVALITIQTN---PMYVNTQASAEVKKLRQ 236
DB 150 RSELEAANQGVNTLRLOQGNATAGQ-AVVALVDQSGYVAYEEETKLPGIRVGMBAQ 208
QY 237 IAEGLLAADVGVIANIGIFEDGTVPPEKGRLLFADPVNVESTGQITLRAAVPNDQ---NI 293
DB 209 V---RLMSGDQPI-----DGT-----VESISSGITDRNSTPDGGLLANV 244
QY 294 -----LMPGLYVRLMDQYAVD 310
DB 245 EPTFWVRLAQRIPVRIRLDQVPAD 269

RESULT 7
US-09-815-242-11146
Sequence 11146, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl U.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11146
LENGTH: 390
TYPE: PRT
ORGANISM: Haemophilus influenzae
US-09-815-242-11146

Query Match 7.8%; Score 156.5; DB 9; Length 390;
Best Local Similarity 21.6%; Pred. No. 9.8e-06;
Matches 71; Conservative 51; Mismatches 115; Indels 91; Gaps 10;

QY 75 VRAOVGGIIQKRLFOEGSYVAGQPLVQIDSTYEANLESARQLATA----- 122
DB 64 VSSQVAGNVAKINDMNDKVHAGDILVELDNTNKLSPQKSNLNAVRQVEQLGFTVQ 123
QY 123 -----QATLAKADADLARYPELVAAEAVSRQEDYDAVATKRSAGYKAAQ----- 168
DB 124 QLOSAVHANEISLAQAGNLARVQLERKGAIDKESFQHAKEAVELAKANLNASKQGLAA 183
QY 169 -----AAKSGINLRSRITAPISGRTGQSKVSEGTILNAGDTT 208
DB 184 NQALLRNPVLEBQIQNAINSLKQAMINLORTIRSIDIVVARNVQVQANVSGAL 243
QY 209 VLAITRQTNPMYVNTQASAEVKKLRQI-----AEGKLLAADGVIAVIGIFEDGTTYPE 263
DB 244 MAVVSNQEMLEANFKETQLTNMRIGQPVKIHFDLYGKKEFDGVIN-GIEMGTGNAF-- 300
QY 264 KGRLLFADPVNVESTGQITLRAAVPNDQNIIMPGLYVRLMDQVAVNDAFVVPQO----- 318
DB 301 --SLI--PSONATGNMVKVQRP-----VRIKLD-----PQGFETTP 334
QY 319 -AVTRGADVTWIVNAQGM-----EPR 340
DB 335 LRIGLSATKVRISDSSGAMLRKTERK 362

RESULT 8
US-09-769-744A-94
Sequence 94, Application US/09769744A
Publication No. US20030134407A1
GENERAL INFORMATION:
APPLICANT: Le Page, Richard W.
APPLICANT: Wells, Jeremy M.
APPLICANT: Hanniffey, Sean B.
APPLICANT: Hansbro, Phillip M.
TITLE OF INVENTION: Proteins
FILE REFERENCE: PWC/P2112WO
CURRENT APPLICATION NUMBER: US/09/769,744A
CURRENT FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: PCT/GB99/02452
PRIOR FILING DATE: 1999-07-27
PRIOR APPLICATION NUMBER: GB 9816336.3
PRIOR FILING DATE: 1998-07-27
PRIOR APPLICATION NUMBER: US 60/125329
PRIOR FILING DATE: 1999-03-19
NUMBER OF SEQ ID NOS: 196
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 94
LENGTH: 399
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-769-744A-94

Query Match 7.5%; Score 151.5; DB 12; Length 399;
Best Local Similarity 24.9%; Pred. No. 2.9e-05;
Matches 98; Conservative 64; Mismatches 163; Indels 69; Gaps 18;

```

QY 12 AALAAVALVLSGCGGGA--AQQGQPRGRAPAVGVATVAPCTVATVLPGLS 69
Db 15 AALGAASVVL---GAGGILLFRQPSQTHAKDEP---THLVAKESGVASVLLSGVTA 68
QY 70 LFTADVRAQV-GITQKRLFQEGSVYRACQPLVQIDSTYEANLIESARQALATA----- 122
Db 69 KNEQYVYPDASKGDLEIILVSVGDKYSEGOALVKYSSSEAAVYDASRAVARADRHINE 128
QY 123 --QATLAKADADLARI-KPLVAAEANSRQEYDAAVTAKGSAEGVYRAQAALISAGINLN 179
Db 129 LNQANNEAASAPAPQIPAPVGGEDATVQSPTPVAGNSVASIDQLGDARPARADAAQLS 188
QY 180 RSRTAPLPGIFGQSKVSEGTLLNAGDTVYLATIRQTNPMYVAVTOS--ASEVM----- 231
Db 189 KA-----QSQDQATVVL---SILEGIVEVNS---NVSKSPITASQOVMMHIVS 230
QY 232 ---KLRRQIAGBKLMAADGVIAVGIKPDDGVYPER---GRLLF-----ADPV 273
Db 231 NENLQVKGELSEYNL--ANLSVGQEVSF-TSKYKPYDKMKTGKLSYSDYKKNNGEAASPA 287
QY 274 VNESTG-QITLRAAVPNDQNIIMPGLYRYVLMDQVAVDNAFVVP-QQAVTRGAKDTVMYI 331
Db 288 AGNNTGSKYPTNIDVTGEVGDCLKQGFSSVNI--EVKSKTALILVPAVSLVMDSDKNYMIY 345
QY 332 NAQGGMEPREVYVAQQQGTNMIYVTSCLKQGDKYV 365
Db 346 DEQQAKAKYEVSLGNADAENQETLSLTNGAKYI 379

```

```

RESULT 9
: US-09-741-669-435
: Sequence 435: Application US/09741669
: Patent No. US20020022718A1
: GENERAL INFORMATION:
: APPLICANT: Forsyth, R. Allyn
: APPLICANT: Ohlsen, Kari L.
: APPLICANT: Zyskind, Judith W.
: TITLE OF INVENTION: Genes identified as required for
: proliferation of E. coli
: FILE REFERENCE: ELITRA.009A
: CURRENT APPLICATION NUMBER: US/09/741,669
: CURRENT FILING DATE: 2000-12-19
: PRIOR APPLICATION NUMBER: US 60/173005
: PRIOR FILING DATE: 1999-12-23
: NUMBER OF SEQ ID NOS: 481
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 435
: LENGTH: 355
: TYPE: prt
: ORGANISM: Escherichia coli
: US-09-741-669-435

```

Query Match	6.9%	Score 138.5	DB 9	Length 355
Best Local Similarity	26.3%	Pred. No. 0.00036		
Matches	52	Conservative	32	Mismatches 55; Indels 59; Gaps 6

QY	65	GRLESLRTADVRAVQGGIIQKRLFOEGSVYRAGQPLYOIDSSTYEANLESARAQLATAQA	124
		: : : : : : : : : : : : : : : : : : : :	
Db	41	GRIEATE-VDIASKIAGRIDTLIVKEGKFRGEQEVLAQKMDTVLQGRULEAIAQIREAGS	99
		: : : : : : : : : : : : : : : : : : : :	
QY	125	TLVADADNL-----AYKRLVAEAISRQED-----	151
		: : : : : : : : : : : : : : : :	
Db	100	AVAAQAQLLEGRQSTTRPAQSLVNORQAELEDSVAKRHTSRSLAQALSAQOQLDDDRRA	159
		: : : : : : : : : : : : : : : :	
QY	152	--AAVTAKRSAGYKAAQAALIKSAGIN-----	188
		: : : : : : : : : : : : : : : : : : : :	
Db	160	AESAPAALESKAQVSASKAATEARTNIIQAQTRVEAAQATERRIADIDDSSELKAPRD	219
		: : : : : : : : : : : : : : : : : : : :	
QY	189	GFIGOSKYSF--CTLINAG	205
		: : : : : : : : : : : : : : : : : : : :	
Db	220	GRV-QYRVAEPGEVLAAG	236
		: : : : : : : : : : : : : : : : : : : :	

```

US-09-988-067B-52      RESULT 10
; Sequence 52, Application US/09988067B
; Publication No. US20030124141A1
; GENERAL INFORMATION:
; APPLICANT: Haas, Rainer
; APPLICANT: Kleantous, Harold
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Miller, Charles
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Odenbreit, Stefan
; APPLICANT: Meyer, Thomas
; TITLE OF INVENTION: Helicobacter Polypeptides and
; TITLE OF INVENTION: Corresponding Polynucleotide Molecules
; FILE REFERENCE: 06132/040002
; CURRENT APPLICATION NUMBER: US/09/988,067B
; CURRENT FILING DATE: 2003-01-31
; PRIOR APPLICATION NUMBER: US '08/831,309
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 52
; LENGTH: 329
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-988-067B-52

```

[illegible]

RESULT 11.
US-08-945-038-4
Sequence 4, Application US/08945038
Publication No. US20020146423v1
GENERAL INFORMATION:
APPLICANT: Doidge, Christopher V.
APPLICANT: Lee, Adrian
APPLICANT: Radcliffe, Fiona J.
APPLICANT: Hocking, Diana M.
APPLICANT: Webb, Elizabeth A.
TITLE OF INVENTION: PROTECTIVE HELICOBLASTER ANTIGENS
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY & LARDNER
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20007-5109
COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/945.038
FILING DATE: 23-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU96/00225
FILING DATE: 19-APR-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PN2575
FILING DATE: 21-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PN3931
FILING DATE: 03-JUL-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PN7565
FILING DATE: 16-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Bent, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 017227/0133
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 329 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-945-038-4

Query Match 6.1%; Score 122.5; DB 8; Length 329;
Best Local Similarity 22.8%; Pred. No. 0.0087;

Matches 63; Conservative 37; Mismatches 109; Indels 67; Gaps 6;

QY 63 LPRGLESRTADVRQVGGIIOKRLFOEGSVYRAGQPIYQIDSSYTYEANLESARA----- 117
DB 38 LQGFLEA-REYSVSSKVRERIEKVFVKGDRIKKDLYFSISSPELEAKLAQAEKGHQA 96
QY 118 -----QATAQATLA-----KADADLAR 135
DB 97 KALSDVVRKSGRDETINSABDVQAAKQATLAKETTKRYQDLYONGVASLQKREAVAA 156
QY 136 YKPLVAABAVSRQETDAV-----TAKSAAEAGVAAQQAISAGINLRSRTITPISGF 190
DB 157 YESTKYNESAAYQKYKMAALGASSESKIAAKAKESAAALGOVNEVESYLKDVATAPIDGE 216
QY 191 IGSQSVSEGLTNAGDTVTYLAIRQTNPMYVNVVQSSASRVMKLRRQIAG----- 240
DB 217 VSNVLISGSELSPKGFVPLMTDLKDSWIKTSVPKYLNDPFGVKEP-BGYIPALKRSK 275
QY 241 ---KLLAADGVAVGIRKFDGTVPYKGRLLFADPV 273
DB 276 FRVKYLSVMGDFATMKATNNSTVYMKGYEVAIPL 311

RESULT 12
US-09-784-208-4

; Sequence 4, Application US/09784208
; Patent No. US20010019836A1

; GENERAL INFORMATION:

; APPLICANT: IZUI, HIROSHI

; APPLICANT: ONO, EIJI

; APPLICANT: MATSUI, KAZUHIKO

; APPLICANT: MORIYA, MIKA

; APPLICANT: ITO, HISAO

; APPLICANT: HARA, YOSHIIKO

; TITLE OF INVENTION: L-GLUTAMIC ACID-PRODUCING BACTERIUM AND METHOD FOR

; FILE REFERENCE: 0010-0989-0

; CURRENT APPLICATION NUMBER: US/09/784.208
; CURRENT FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: 09/271,438
; PRIOR FILING DATE: 1999-03-18
; PRIOR APPLICATION NUMBER: JP 10-69068
; PRIOR FILING DATE: 1998-03-18
; PRIOR APPLICATION NUMBER: JP 10-297129
; PRIOR FILING DATE: 1998-10-19
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 4
; LENGTH: 407
; TYPE: PRT
; ORGANISM: Enterobacter agglomerans
US-09-784-208-4

Query Match 6.1%; Score 122.5; DB 9; Length 407;

Best Local Similarity 20.9%; Pred. No. 0.012;
Matches 83; Conservative 58; Mismatches 154; Indels 103; Gaps 13;

QY 15 AAVALVSSCGKGDAAGGQPAAGREAPAVGVTVHPQTVALTVELPGRLESRTAD 74
DB 14 SVADATVATWHKKKGDAVSDE-----VIVEITDKVLEVP----- 50
QY 75 VRAQVGGIIOKRLFOEGSVYRAGQPIYQI-----DSTYEANLESARAQATLAQATL 126
DB 51 --ASADGVLEAVLDEGATVTSRQILGRLKEGNSAKGESSAKASNDTTPQRORTASLEE 108
QY 127 AKADA-----DLARYKPLVAABAVSRQETDAVAVTAKSAAEAGVAAQQAIAIK 172
DB 109 ESSDALSPAIRRLIAEHNLDAAQIKGTGVGRLTRDEVEKHLAKPQAEKAAAPAGAT 168
QY 173 SAGINLNRSRITAPISGFIGQSKVSEGLTNAGDTVTYLAIRQTNPMYVNVVQSSASEVMK 232
DB 169 AQGVANRSEKRVEMTRL--RKVAERLEKXNSTAMLTNNEIN-----MKPIMD 217
QY 233 LRQIAE-----GKLLAADGVAVGIRK-FDDGTVPYKGRLLFADPVNESTGQITLR 284
DB 218 LRQYGDAPFERKHGVRLL---GFMSFYIKAVAEALKRYPEVNASIDGEDVYVHNYFDVISA 274
QY 285 AAVENDQNIIMPGLYRVLMQVAVDNAFVVPQAVTRGADVTWIVAAQGMREBYTV 344
DB 275 VSTR-----GLVTPLRDVDALSMADI-----EKKIKELAVKGRDG---KLTV 315
QY 345 AQOQTNWIVTSGLKDGKVVVEGISTAGTGAKKTP 382
DB 316 DDLTGNNFTTNG-----GVGSLMSIP 338

RESULT 13
US-10-078-107-2

; Sequence 2, Application US/10078107
; Publication No. US20020182688A1

; GENERAL INFORMATION:

; APPLICANT: IZUI, HIROSHI

; APPLICANT: HARA, YOSHIIKO

; APPLICANT: SATO, MASAKAZU

; APPLICANT: AKIYOSHI, NAOKI

; TITLE OF INVENTION: METHOD FOR PRODUCING L-GLUTAMIC ACID

; FILE REFERENCE: 219846USO

; CURRENT APPLICATION NUMBER: US/10/078.107

; CURRENT FILING DATE: 2002-05-13

; PRIOR APPLICATION NUMBER: JP 2001-044134

; PRIOR FILING DATE: 2001-02-20

; NUMBER OF SEQ ID NOS: 8

; SOFTWARE: Patentin version 3.1

; SEQ ID NO 2

; LENGTH: 407

; TYPE: PRT

; ORGANISM: Enterobacter agglomerans

US-10-078-107-2

Query Match 6.1%; Score 122.5; DB 14; Length 407;

	Best Local Similarity	20.9%;	Pred. No. 0.012;	
	Matches	83;	Conservative	58; Mismatches 154; Indels 103; Gaps 13.
QY	15	AAAVLVLS	CGCGDAAQGGG	PAGREAPAVGVVTVHPTQVATVVEIPGRLESTRYAD 74
Db	14	SVADATVATW	HKKPPGDAVS	VRDE-----VIVEIEDKVLVEP----- 50
QY	75	VRAQVGII	IQKRLFEQGS	TVRAQRLYQI-----DSSTYEANLESARQQLATAQTL 126
Db	51	--ASADGVLE	AVLEBEGATVTSRQ	ILGRLKEGNSAGKESKASSAENSDTTPAQORTASLEE 108
QY	127	AKADA-----	DLARYPELVAAE	NSQGEVDAVTVARSGAEAGVKAQAQAIK 172
Db	109	ESSDLSPAIR	RLLAENHLMDAAQ	IKGTGVGRLTFREDEKHLAMPQAEKAAPAAAGYAT 168
QY	173	SAGINLNSR	ITAPISCFIGOS	SEGLINAGDPTVATIRQTPMRYVNTQSSASEYWK 232
Db	169	AQGPVANS	SEKRVFPTRL--	RKRAVERLLEKANSIMLTTEIN-----MKPLMD 217
QY	233	LRQQLAE----	GKLLAADGVIA	IGIK-FDDGTVPYSEKGRLLFADPVNVESTGQITLR 284
Db	218	LRQUGDAF	EKRHHGVRLL--	GFMSFYIAAVEALIKRYPEVNASIDGEDVVVYHNFEDVIA 274
QY	285	AAVPRDQNI	LMPLGLYVRLM	QVAVDAAFFVVPQQAIVTGAQDTWIVVAQGSMEPREYTV 344
Db	275	VSTRP-----	GLVTPVLRD	VDVALSMADI-----EKIKELAVVAGRDG-----KLTV 315
QY	345	AQGGGTWIV	TSGLKDGD	KVVVEGSIINGITGAKKVT 382
Db	316	DDLGTG	NFTTNG-----	GVFGSLMSTP 338

```

RESULT 14
US-10-077-751-2
: Sequence 2, Application US/10077751
: Publication No. US20020192772A1
: GENERAL INFORMATION:
: APPLICANT: SATO, MASAKAZU
: APPLICANT: AKIYOSHI, NAOKI
: TITLE OF INVENTION: METHOD FOR PRODUCING L-GLUTAMIC ACID
: FILE REFERENCE: 219849US0
: CURRENT APPLICATION NUMBER: US/10/077,751
: CURRENT FILING DATE: 2002-05-13
: PRIOR APPLICATION NUMBER: JP 2001-044135
: PRIOR FILING DATE: 2001-02-20
: NUMBER OF SEQ ID NOS: 8
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 2
: LENGTH: 407
: TYPE: PKT
: ORGANISM: Enterobacter agglomerans
US-10-077-751-2

```

Query Match	6.1%;	Score 122.5;	DB 14;	Length 407;
Best Local Similarity	20.9%;	Pred. No. 0.012;		
Matches	83;	Conservative	58;	Mismatches 154;
				Indels 103;
				Gaps 13;
QY	15	AAAVLVLSGCGKGDAAQGGOPAGREAPAEVGVVTHPQTVALTYELPGRLSRLTAD	74	
DB	14	SVAADATVAWKKKGGDAVSDE-----VIVEITDKVLEVP-----	50	
QY	75	VRAQVGGIIQRLTPOEGSYVAGQPLVQI-----DSSTYEANLESARQLATAQATL	126	
DB	51	--ASADGLEVLEDEGATVTSRQILGRLKEGNSAGKSSKKAESNDTPPQROTALEE	108	
QY	127	AKADA-----DLARYKPLVAEAEVROEYDAVTAATKRSAAEGVKAQAAIK	172	
DB	109	ESSDALSPAIRRLIAEHNLDAQIKGCGVGRLTREDVEKELANKPQLEKAAAPAAGAT	166	
QY	173	SAGINLNRSRITAIISGFIGSKYSEGITLNAQDTVLATIRQINPMRYVNTQASAEVWK	232	
DB	169	AQQFVANKSERKVPMTRL--RKRAVAERLLEKNSNTALTYTFNEIN-----MKDIMD	217	

```

QY      233 LRQJIAE-----GKLLADGVIAIGIK--PDDGTVYPEKGKLLFPADPVVNESTGQITLR 284
Db      218 LRQJGADPEKRRHGVRL---GMSFYIAKVAEALKRYPEVNASIDGEDVYVHYHPFVSIA 274
QY      285 AAVPEQDQNIIMPGLVYRVLMDQVAADNAFVVVPGQAVTRBAKDTVMIVNNAQGMREPREVTV 344
Db      275 VSTPR-----GLVTPVLRVDVALISMADI-----EKKIKELAVKGRDG---KLTV 315
QY      345 AQQQSTNMIIVTSGLKDGDKRVVVEGISIAGITGAKKVT 382
Db      316 DDLTGNGFTITNG-----GVFSGSLMSTP 338

```

```

RESULT 15
US-10-315-023-4
; Sequence 4, Application US/10315023
; Publication NO. US20030119153A1
; GENERAL INFORMATION:
; APPLICANT: IZUI, Hiroschi
; APPLICANT: ONO, Ei-ji
; APPLICANT: MATSUI, Kazuhiko
; APPLICANT: MORIYA, Mika
; APPLICANT: ITO, Hisao
; APPLICANT: HARA, Yoshihiko
; TITLE OF INVENTION: L-GLUTAMIC ACID-PRODUCING BACTERIUM AND METHOD FOR PRODUCING L-GLU
; TITLE OF INVENTION: ACID
; FILE REFERENCE: 0010-0989-0
; CURRENT APPLICATION NUMBER: US/10/315, 023
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: US/99/271, 438
; PRIOR FILING DATE: 1999-03-18
; PRIOR APPLICATION NUMBER: JP10-69068
; PRIOR FILING DATE: 1998-03-18
; PRIOR APPLICATION NUMBER: JP10-297129
; PRIOR FILING DATE: 1998-10-19
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 407
; TYPE: PRT
; ORGANISM: Enterobacter agglomerans
US-10-315-023-4

```

Query Match	6.1%	Score 122.5	DB 15	Length 407
Similarity	20.9%	Pred. No. 0.012		
Matches	83	Conservative 58	Mismatches 154	Indels 103
			Gaps 13	
Oy	AAAVALVSSCGKGDAAAGGAPAGRAPAPVVGVTVPHPQTVALLTELPGRLESLRTAD	74		
Db	SVADATVATMHHKRGDVAVSDE-----VIVEIETDKVLEVP-----	50		
Oy	VRAQVGIIQKRLFOEGSSYVRAGPLVYI-----DSSTYEANLESARQIATAQATL	126		
Db	--ASADGVLAEVLEDEGATVTSRIILGRLEKGNAGKRESSAKAESNDTTPAQRTASILEE	108		
Oy	AKADA-----DIARYKPLVAABANSRQETDAATVAKRSALRGVKAQAQAIIK	172		
Db	ESSALSPAIRRLIAEHNLDAAQIKGTGVGGRILREDBVEKHANKPOAEKKAAPAAAPAAAT	168		
Oy	SAGINLMSRRTIAPISGFIGOSKXSECTLLNAGDTVLATIRQTNPMYVNTQASASVMK	232		
Db	AQGPVARSERKRVPMTRL--RKRAVEHLLAKNSTAMLTTFENEIN-----MKDIMD	217		
Oy	LRROIAE-----GKLLAADGVIAVGK--FDDDTVPYPERGKRLFFADPVVNESGTQITLR	284		
Db	LRKQYGAPEKRGHVRL---GFSMFYIKAAVEALKRPEVNAASIDGEDVYTHNFVDSIA	274		
Oy	AAVENDONILMPGLYYRVLMDOVAVDNAFVVPQAAVTRGAKDPTMIVNAQSGMPEEVT	344		
Db	VSTR-----GLVTFVLRDVALSNADL-----EKKIKELAVKGRDG---KLTV	315		
Oy	AQOQGTWVITVSGLKGDGKVVVEGISIAGITGAKKPTP	382		

Db 316 DDLTGNGFTITNG-----GVFGSLMSTP 338

RESULT 16

US-10-315-023-10

Sequence 10, Application US/10315023

Publication No. US20030119153A1

GENERAL INFORMATION:

APPLICANT: IZUI, Hiroshi

APPLICANT: ONO, Eiji

APPLICANT: MATSUI, Kazuhiko

APPLICANT: MORIYA, Mika

APPLICANT: ITO, Hisao

APPLICANT: HARA, Yoshihiko

TITLE OF INVENTION: L-GLUTAMIC ACID-PRODUCING BACTERIUM AND METHOD FOR PRODUCING L-GLUTAMIC ACID

TITLE OF INVENTION: ACID

FILE REFERENCE: 0010-0989-0

CURRENT APPLICATION NUMBER: US/10/315,023

PRIOR FILING DATE: 2002-12-10

PRIOR APPLICATION NUMBER: US/09/271,438

PRIOR FILING DATE: 1999-03-18

PRIOR APPLICATION NUMBER: JP10-69068

PRIOR FILING DATE: 1998-03-18

PRIOR APPLICATION NUMBER: JP10-297129

PRIOR FILING DATE: 1998-10-19

NUMBER OF SEQ ID NOS: 15

SOFTWARE: PatentIn version 3.1

SEQ ID NO 10

LENGTH: 407

TYPE: PRT

ORGANISM: Enterobacter agglomerans

US-10-315-023-10

Query Match

Best Local Similarity 6.1%; Score 122.5; DB 15; Length 407;

Best Local Similarity 20.9%; Pred. No. 0.012;

Matches 83; Conservative 58; Mismatches 154; Indels 103; Gaps 13;

Db 15 AAVALVSSCGGDDAAGGQAPGRAPAPVGVVTPVTPVLTVELPGLSLRAD 74

14 SVADATVATWKKKPPDDAVSRDE-----VYIEIETDKVLEVP----- 50

Qy 75 VRAVGGIITQKRLFOEGSVYRAGQPLYOI-----DSSTYEANLESARQATATL 126

Db 51 --ASADGVLEAVLEDEGATVTSRQLTGRLEKNSAGKSSAKAESNDTPPAQRTASLE 108

Qy 127 AKADA-----DLARYPLVAAEVNSQEDDAVTAARSAAGYKAQAALX 172

Db 109 ESSDLSPAIRRLIAEHNLDAQIKGTGVGRLTREDEKHLANKPOAEKAAPAGAT 168

Qy 173 SAGIMLNSRITAPISGFIGSKVSEGLTLNAGDTTVALTROTNPMTVNTQSASEYWK 232

Db 169 AQPAPVANSSEKVPVTRL--RKVAERLLEAKNSTAMLTTFEIN-----MKPIMD 217

Qy 223 LRRQIAE-----GKLLADGVIAVGIR--FDDGVTPYEPKGRLLFADPVVNESTGQITLR 284

Db 218 LRKQYGDAPFEKHGVRLL--GFMSFYIKAVVEALRKRYEVNASIGDEVDVYHNPVDSIA 274

Qy 285 AAVPDQNILMGLYVRVILMDQVADNAFVVPQAVTGAKTQWIVVNAOGEMERETV 344

Db 275 VSTPR-----GLVTPVLRVDVDAISMADI-----EKIKELAVKGRDG---KLTV 315

Qy 345 AAOQGTNMTVTSGLKDGKVVVEGISIAGITGAKKVT 382

Db 316 DDLTGNGFTITNG-----GVFGSLMSTP 338

RESULT 17

US-10-193-764-61

Sequence 61, Application US/10193764

Publication No. US20030133943A1

GENERAL INFORMATION:

APPLICANT: Loosmore, Sheena M.

APPLICANT: Yang, Yan-Ping

APPLICANT: Klein, Michel H.

TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH MOLECULAR WEIGHT PROTEINS

FILE REFERENCE: 1038-1239MIS

CURRENT FILING DATE: 2002-07-12

PRIOR APPLICATION NUMBER: 09/167,568

PRIOR FILING DATE: 1998-10-07

NUMBER OF SEQ ID NOS: 91

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 61

LENGTH: 1180

TYPE: PRT

ORGANISM: Haemophilus influenzae

US-10-193-764-61

Query Match

Best Local Similarity 6.0%; Score 120.5; DB 12; Length 1180;

Best Local Similarity 20.7%; Pred. No. 0.088;

Matches 89; Conservative 66; Mismatches 178; Indels 97; Gaps 18;

Qy 28 GDDAAGGQAPGRAPAPVGVVTPVTPVLTVELPGR-----LESRTADVRAQVGII 83

Db 750 GGDVLAVGNISG-----NAVVTANSGLTT-LAGSTIKGTESITTSQSGNIGKI 800

Qy 84 QKRLFOEGSVYRAGQPL-----YQIDSTYEANLESARQATATLAKADADLARYKP 138

Db 801 SKGTVN-----VKATNSLTTOADSKIEATEGEANTSKTSIIG--GTISGTVETVATEG 853

Qy 139 LVAAEAVSRQEDYDAVTAARSAENG-----VKAQAALISAGINLNSR 182

Db 854 LTTQAGSTITGTESVTSQSGNIGMISGKVEVSATKDLITSGSEIKATAGEVNTS 913

Qy 183 ITAPISGFIGSKVSEGLTLNAGDTV-----LATIRQTNPMVNVVNTQSASEYWK 231

Db 914 ATGITDITGSGNTVN--VTATGDLTVEDAKIDATGCAATLITSGKL--TTAASSIT 969

Qy 232 KLRQIAEGKLLADGVIAVGIRFD-----GTVEPKGRLLFADP---VVN---EST 278

Db 970 SANNOV--NLSAKDGSIGGINANVTLTNTGALTTVKSSINANSGLVINKADELN 1026

Qy 279 GQITLRAVPNDQILMGLYVRVILMDQVADNAFVVPQAVTGADTQWIVVNAOGME 338

Db 1027 GEASGNHTVNVATANGSGSVIATTSRVNI-----TGDLITING--- 1066

Qy 339 PREVTAAOQGTNMTVTSGLKDGKVVVEGI-SIAGITGAKKVPK-EMASSENOAAPQ 396

Db 1067 ---LNTISKGINVTLKGVKIDVKYIQPGIASVDEVIEAKRIEKKVLDSDEREALAK 1123

Qy 397 SGVQTASEAK 406

Db 1124 LGVSAVRFAE 1133

RESULT 18

US-10-193-764-59

Sequence 59, Application US/10193764

Publication No. US20030133943A1

GENERAL INFORMATION:

APPLICANT: Loosmore, Sheena M.

APPLICANT: Yang, Yan-Ping

APPLICANT: Klein, Michel H.

TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH MOLECULAR WEIGHT PROTEINS

FILE REFERENCE: 1038-1239MIS

CURRENT FILING DATE: 2002-07-12

PRIOR APPLICATION NUMBER: 09/167,568

PRIOR FILING DATE: 1998-10-07

NUMBER OF SEQ ID NOS: 91

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 59

LENGTH: 1188

TYPE: PRT

ORGANISM: Haemophilus influenzae
US-10-193-764-59

Query Match 6.0%; Score 120.5; DB 12; Length 1188;
Best Local Similarity 20.7%; Pred. No. 0.089;
Matches 89; Conservative 66; Mismatches 178; Indels 97; Gaps 18;

QY 28 GDDAAGGQOPAGREAPAVGVVTVHPQVALTYELPGR-----LESRTADVRAQVGII 83
DB 758 GGDTLAVNISG-----NAVTVANSGLTT-LAGSTIKGTEISITSSQSGNGIKI 808
QY 84 QKRLFOESSYRACQPL-----YQDSSTEANLESAPAQLATAQATLAKADADLARKP 138
DB 809 SGKTVN---VKATNSLTQADSKIEATEGEANVTSKTSIIQ---GTSIGGTVEVATEG 861
QY 139 LVAEAVSRQREYDAVAVTAARSAEAG-----VKAQAQIKSAGINLNSR 182
DB 862 LTTQAGSITTTGTESTTSQSGNIGMISGKVEVSATKDLITSSGSIKATAGEVNTS 921
QY 183 ITAPISGFIGOSKVSSEGTLLNAGDTTV-----LATIRQTNPMYVNTQSASEVM 231
DB 922 ATGTIDGITSNTVNV--VTANTGDLTVEDA KIDATGAGATLLTATSGL--TTKASSSIT 977
QY 232 KLRQIAGSKLLADGVAVGKFPD-----GTVPKGRLLFPDP---VNV---EST 278
DB 978 SANNOV---NLISAKDGISGININAMVTLNTTGALLTVKSSINANSGLTVINAKDAELN 1034
QY 279 GQITLRAVPNDQNLMPGLVVRVLMDOVADNAFVVQAVTGAKQDTWIVAAQGME 338
DB 1035 GEASGNTVAVATNANGSGSVIATTSRVNI-----TGDLITING--- 1074
QY 339 PREYVAOQOGTNNIVTSGLDKDGKVVVEGI-SIAGITGAKKVTPK-EMASSENQAAAPQ 396
DB 1075 ---LNIISKNGINTVTLKGVNIDVYIQTGASVDEVIEARKILEKVXKDSDEERELAK 1131
QY 397 SGVOTASPAK 406
DB 1132 LGVSAVRPAE 1141

RESULT 19

US-10-156-761-9338
Sequence 9338, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMIIDA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156, 761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 9338
LENGTH: 696
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-9338

Query Match 5.9%; Score 119.5; DB 15; Length 696;
Best Local Similarity 22.0%; Pred. No. 0.049;
Matches 97; Conservative 63; Mismatches 162; Indels 119; Gaps 20;
QY 52 VHPQVALTYELP-----GRLESRTADVRAQV-GGIIOKRLFOESSYVRAG 97
DB 1132 LGVSAVRPAE 1141

DB 161 LHP--APLVQPLPIGTODDFRGVVDLVNMSLLMADGRDIVEGSPVDPALRDEAD--RRR 216
QY 98 QPLVQDSSTEANLESAPAQLATAQATLAKADADLAR-----YKPLV 140
DB 217 RQLEEAVALHPLVALEEFCAQATYSQTLTRALDLRTGTGGVVVLCGSAVRNKGIEFL 276
QY 141 AEAIVSRQREYDAVTAARSAEAGVKAQAQIKSAGINLNSRITAPISGFIGOSKVSSEGT 200
DB 277 AA-AVALVPSPLDPAVRGTEDGAVRBRADPAAPFAALVFKVATATGRLTYLVYSGT 335
QY 201 LMANQDTTVLATIRQTNPMYVNTQSASEVMKLRQIABEKLTA----- 244
DB 336 -IGKEFTVDTGAARSSERIRILRVQDRHTELDRAVA-GDIVAVGPKTARAGSTLCAP 393
QY 245 -----ADGVAVGKFPDGTVPKGRLLFPDPVNVNBSGQITLRAAVPNDQNI 293
DB 394 AAPLVLEPPTADVAVVAEARSST---PTDRLASLALAEEDPSLVVTRTDEPTQTV 450
QY 294 L--MPGLYVRVLMDOVADNAFVVP-----QQAIVTGAKQDTV----- 328
DB 451 LSGWGEHLHEVAVVEKIRDRGTAVNVGRPRVAVRETVARGVSGLVYRHKQDGGAGQFAH 510
QY 329 MIVNAQ-----CGMPEYTVAAQOGTNNI--VTSGLDKDGKVVVEGSIAG--I 374
DB 511 VVLDVEPLESAADGCHGFERSAVGGRVQEVYRAVAAGCRD---ALAEQ-PLGGHPV 566
QY 375 TGAKKVT-----PKEMASSE 389
DB 567 TGLRVVLTGATHPKD---SSS 585

RESULT 20

US-10-315-023-11
Sequence 11, Application US/10315023
Publication No. US20030119153A1
GENERAL INFORMATION:
APPLICANT: IZUI, Hiroshi
APPLICANT: ONO, Eiichi
APPLICANT: MATSUI, Kazuhiko
APPLICANT: MORIYA, Mika
APPLICANT: ITO, Hisao
APPLICANT: HARA, Yoshihiko
TITLE OF INVENTION: L-GLUTAMIC ACID-PRODUCING BACTERIUM AND METHOD FOR PRODUCING L-GL
TITLE OF INVENTION: ACID
FILE REFERENCE: 0010-0989-0
CURRENT APPLICATION NUMBER: US/10/315, 023
CURRENT FILING DATE: 2002-12-10
PRIOR APPLICATION NUMBER: US/09/271, 438
PRIOR FILING DATE: 1999-03-18
PRIOR APPLICATION NUMBER: JP10-69068
PRIOR FILING DATE: 1998-03-18
PRIOR APPLICATION NUMBER: JP10-297129
PRIOR FILING DATE: 1998-10-19
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn version 3.1
SEQ ID NO 11
LENGTH: 407
TYPE: PRT
ORGANISM: Escherichia coli
US-10-315-023-11

Query Match 5.9%; Score 118.5; DB 15; Length 407;
Best Local Similarity 21.4%; Pred. No. 0.027;
Matches 86; Conservative 55; Mismatches 151; Indels 109; Gaps 15;
QY 15 AAVALVLSGCGGDAAGGQOPAGREAPAVGVVTVHPQVALTYELPGRLESRTAD 74
DB 14 SVADATVATVHKKRGDAV-----VRDEVLEIETDKVVLVEVP----- 50
QY 75 VRAQVGIIQKRLFOESSYRACQPLVQI---DSSTEANLES-----ARAQLATAQ- 123
DB 51 --ASADQILDVAVLEDEGTVTYSRQIIGRLREGNSAKETSAKSEKASTPAORQOASLEE 108

124 -----ATLAKADADLARYKPLVAEAVSROEYDAAVTAKRSAAEAGVKAQAIAIK 172
109 ONNDALSPAIRRLIAEHMIDAAQIKGTGVGRLTREVDEKHLANKPOAEKAAAPAGAT 168
173 SAGINLNSRITAPISGIGOSKSEGTULNAGDTTVALTIROTPMYVNTQASAEWK 232
169 AQPVAANSERKVPMTIR--RRKVAERLLEAKNSTAMLTTFEVA--MKPIMD 217
233 LRQIAEKLAAADGVIAGIKFDGTV-----YPERGRLIFADPVVNESTQITLRBA 286
218 LRQVGEA-FEKRHGRIRLGFMSFYKAVVEALKRYPEVNASIDGDDVYVNTQASAEWK 276
287 VPNDONILMPLGLYVAVLMDQVAVDPAFVVPQOAVTRGAKD-----TWIIVAAQGMREPRE 341
277 TPR-----GLVTFVLRD--VD-----TLGMADIEKKIKELAVKGRDG---K 312
342 VTVAAQOGTNNIVTSGLDGDKRVVEGISIAGITAKKVT 382
313 LTVEDLTGNGNFTTNG-----GVFGSLMSTP 338

RESULT 21

US-10-156-761-12596
; Sequence 12596, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT FILING DATE: 2002-05-29
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12596
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-12596

Query Match 5.8%; Score 118; DB 15; Length 382;

Best Local Similarity 23.4%; Pred. No. 0.028;

Matches 97; Conservative 39; Mismatches 171; Indels 108; Gaps 15;
32 AAGGCG-----PAGR--EAPAPVGVVTVHPQVATLVEL-----PGRLESIRADVR- 76
6 AAGSGDNRPTEIPARMBEPPEGPVVLIDQTRLFAEEVELCTDAPALVEAIRTLARVG 65
77 AOVGSIQKRLFOEBSY-----VRAGPFLYQIDSTYEANLESAAQATQA 124
66 APLUCI-----AGAYVLAALAAANGFVDEKAQALSGARPTAVNLAVGRRARAAAYDA 118
125 TLAKADADLARYKPLVAEAVSROEYDAVTAKRSAEAGVKAQAIAIKSAGINLNSRIT 184
119 LAGGDPERRAAALQAARALHREDAEASA--RMAEHGLALDELPCGGHRL----- 169
185 APISGFIOGSKVSEGTULNAGDTTVALT-----IRCTNPMYVNTQASAEV 230
170 -----LTHCNTALVSGEGTAFAYALAAREGRRLRWVDTRPLLOGARLTAYEA 221
221 MK-----LRQIAEKLAAADGVIAGIKFDGTVYVEKGRLLFADPVVNESTQITLR 284
222 ARNGCAVTLTIDNAGSLFAAGEVDVAVLIGAD-----RIADDSVANKVGSYPL- 270
285 AAVPNDONILMPLGLYVAVLMDQVAVDPAFVVPQOAVTRGAKDTVMI-VNAAQGMREPREV 343

271 -----AVLARHHPPIVAVPTTVDPDPDGSIEBQRAHEVTEV 314
344 VAAQOGTNNIVTSGLDGDKRVVEGISIAGITAKKVTPEKMA--SENGAAP 395
315 APQ-----VFVAGAEAGGIPVAPLGTQAVNPAFVDVTPBELVATVTEBGAVP 363

RESULT 22

US-09-971-536-68
; Sequence 68, Application US/09971536
; Patent No. US20020159976A1
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Hayukkala, Ilkka
; APPLICANT: Lubbers, Leonard
; APPLICANT: Dekker, James
; APPLICANT: Christensen, Anna
; APPLICANT: Holland, Ross
; APPLICANT: O'Toole, Paul
; APPLICANT: Reid, Julian
; APPLICANT: Coolbear, Timothy
; TITLE OF INVENTION: Lactobacillus rhamnosus Polynucleotides, Polypeptides and Methods
; FILE REFERENCE: 1043c2
; CURRENT FILING DATE: 2001-10-02
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/634,238
; PRIOR FILING DATE: 2000-08-08
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/724,623
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: PCT/NZ01/00160
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 68
; LENGTH: 1741
; TYPE: PRT
; ORGANISM: Lactobacillus rhamnosus
US-09-971-536-68

Query Match 5.8%; Score 118; DB 10; Length 1741;

Best Local Similarity 19.6%; Pred. No. 0.26;

Matches 106; Conservative 75; Mismatches 201; Indels 158; Gaps 22;
22 LSSCGKGDAAGGQAPGRAP--APVVGVTVPQVVA--LYVELPRLSIRAD 74
388 LSDGAKKLAQAATNNYQLTADDLAKVTGATITTPATTSVDSDVSFEYDKTKASEAAG 447
75 VRAGV-----GGIIQ--KRLFOEGSYRAGQPLVQIDSTYEANLES----- 115
448 IQATIKLDTGKVVLTAADIIVTNDVNAAGYSYL--SDGAKKLAQAATGNNYQLTADDL 506
116 -----RAQLATQAATLAKADADLARYKPLVAE--AVSR 147
507 AKVAGTITTPAVTTVDSDVSFEYDKTKASEAKGQATIKLDTGKVVLTAADIIVTN 566
148 OEYDAATYAKSAAAGVKAQAIA-----ISAG-INLNSRITAPISG-----F 190
567 DDVNAAGVSYLSGAKKLAQAATGNNYQLTADDLAKVMTGITTPPAVNTADSNDLSFEY 626
191 IGOSKVSSEG--TLNAGDTTVALTIRCTNPMYVNTQSA-----SEVMKLRQIAE 239
627 DCKTKASEAKIQAMVNLGTEKTEVDLTSADIVANDVNAAGVSYLSGDKAKKLAQAAT 686
240 GK--LLAADGVAVG-----IKFD-DGVYPERGRLLFADPVVNESTQ 280
687 GNNYQLTADDLAKVAGTITTPATTSVDSDVSFEYDKTKASEAKGQATIKLGEIEKT 746
281 ITLRBA--VPNDONILMPLGLYVAVLMD-----QVAVDNAAF-----VVP 316
747 VDLSSADIIVANDVIV--GKYTSLSDSGSKLQAATGANNYQLTTEVLDRKVSIGSITTP 804

QY 317 QOATRG-----AKTVIMVWNAOGGMEPELVVAQQGCTWIVYSG-----357
 Db 805 AGAATATSKAHPEYDOKTKASEKGTQALITIDIGNEKIVDLTAADIVAEAGVDAGKXY 864
 QY 358 -LDDGKRYVE-----GISLNGTGAKCTPKEMASENQAAPQSGVOTASEAK 406
 Db 865 RLSDDGSKTKOREAGSDHQLTADAEVGTITTPALATDSNDVSEYNGTKRASEAE 924

RESULT 23

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

	Query Match	5.7%;	Score 116;	DB 11,	Length 607;
	Best Local Similarity	24.1%;	Pred. NO.0.083;		
	Matches	71;	Conservative 42;	Mismatches 112;	Indels 70; Gaps 15
QY	ATPAQTATLAKADADLARYPELVAAEVSROEYDAVAVTAKSAENAGVKA--QAAIKSAGIN	177			
	: : : : :				
Db	12 ATTEVALA---SDYEQARAFVASGIVA----TTGGKGRDRIINAGTLALAQALAKTPWMS	64			
	: : : : :				
QY	178 LNRGR-----TPIPGISGTGSKVSEGTLLNAGDTTATLRQINPMYVNTQSASEVMKL	233			
	: : : : :				
Db	65 SDVSRIYLNENAPVIGDVAMETITE-TIIT-ESTMIGHNPQT-PGVGVY--GVGFTIAL	118			
	: : : : :				
QY	234 RR-----QIEGKGLAAD-----GVLAVGIKPEDPDGVNYPEKKR	266			
	: : : : :				
Db	119 GRATLPAAQVYAEIGMTVLIDDAVDFLDAVWMLEALDRGINNVAAIIKKDKGVLY--NNR	176			
	: : : : :				

Qy	267	LEFADPVDNESTGQITLRAAVEND----	ONILMPYLRYRLMDQVADANAF-----	313
Db	177	LKRTLPVVD-----	VTLEQVPEGVMAAEVAAPQVVRILSNPYGIATFFGLSPSEETQA	232
Qy	314	VVYQQAIVTGAKDTWIAVVAAGGMPREVTYAQQGQNTWITVTSGLKQDGKRVYNG		368
Db	233	IYIAPALIGNSAVVLKTPQGDVDVSRKPIYAG-----	NLYISGKKRGKQADVAEG	282

RESULT 24

```

US-10-156-761-13234
Sequence 13234, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMCRA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-212697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 13234
LENGTH: 340
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-13234

```

```

Query Match      5.7%; Score 115; Length 340;
Best Local Similarity 21.8%; Pred. No. 0.043;
Matches 87; Conservative 56; Mismatches 129; Indels 128; Gaps 19;

QY      9 MRAAALAAVALVLSGCGKGDAAGQAPAGREAPAEVGVTVHPQTVALTELPGRL 68
      12 LRAAAVTVGAALTLTLTACGSGSGSGSGSGSG--SGSVKVGILIT-----KTDINPFVK 62
QY      69 SLRTADVRAOVGGIIGKRLFQEGSVYFAGOPLIQIDSTTEANLESAPADLTQATLAK 128
      73 MKEGAETKAAKENGV-----QUSTA---AGK 84
QY      129 ADDDLARFKRLVAAEAENSRQETDAAYTA-KRSAEAGYKAQAQAKSGININRSRITAPI 187
      85 FDDDNA-----GQVTAIENMVAAGVK-----GLITPDSKAIIV 118
QY      188 SGFIGOSKVEGTLINAGDTTVALATTROTNPVNVTVQASBVKLRROIAE-----239
      119 PAL-OKRAKKGVLVALDT-----PTPEBSAVNDLPTIDNLKAGQLIGEYKAYWK 169
QY      240 --GKLLAADGV--IAGGIK---FDDGTVYPEKGRLLFADPVNNESTGQITLRAAVPNDQ 291
      170 KKAKIALDLAPGVSGVGRHNGFLKGFATDRDVCADQGTGGQAKGQITAM-----E 222
QY      292 NIILM--PGLYVRFLMDQVAVDNAFVVPQOAVTGAQKDTNVIINAAG-----MEPRV 343
      223 NCLOKAPGINVVYVITINEPALIGAYTLK--AKGREQDLIVISVDGCGCTQAVAKDKIA 279
QY      344 VAAQQ-----GTNWIVTSGLKDGGKVVVEGSIAGIT 375
      280 ATSQQYPLKMAAGVKAIVTYA-KDGKK--ASGYTDTGVT 316

RESULT 25
US-10-156-761-13251
Sequence 13251, Application US/10156761
Publication No. US20030119018A1

```

GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 13251
LENGTH: 1208
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-13251

Query Match 5.7%; Score 114.5; DB 15; Length 1208;
Best Local Similarity 21.7%; Pred. No. 0.31;
Matches 95; Conservative 53; Mismatches 170; Indels 119; Gaps 21;

QY 11 AALAAALVLSGCGKGDAAAGQAPAGREAPAVGVVYHPVTVALTYELPRLLESL 70
DB 10 AALLSAGVIFALLPVGRAA-ARDAGTPSPPOASTARTVTLVGTDTVT- DLPGGKKTIV 67
QY 71 -----RTADYRAOV-----GGIIOKRLFQSGSVYRAGQPLVQIDSS 106
DB 68 TVERRGATGAVRTQVGVGASVVDLEALPYLAGTDLRLFDVSAALRQGIS- DRK 123
QY 107 TYEANL-----ESARAQI-ATAQATL-----AKADDLAR-YKPLVAEAVSRQSYD 151
DB 124 TGEFLIYGVGKGAARVPSGATRRSLTSGAAVADADKSTFWRSWTRQGIHWMID 183
QY 152 AAV---TKRKAEGAVKAA-----QAIKAGININRSRTAIS- GPIGQSK 195
DB 184 ARVEADMESNAQIGTRAMADGLTGDEVTVAVDLTGDTTHPDLAGRVSRKSTIDEE 243
QY 196 VSEGLTNAGDPTVLATL-----ROTNPMYVAVTQASAEVVK-----LRQIAE 239
DB 244 VAD---RNGHGHVTSYVGGSGAASDGTGRGAP---GATLAVGVILSDQAGSESQIIA 297
QY 240 GKLLAADGVIAVGIKFDDGTVPKGRLLFADPV--VNESTGQITLRAAVPNDQNIIMP 297
DB 298 GMEWAAARVRAIRIVMSIGSTEASDGTDPMAEAVDTLSEETG-----A 340
QY 298 LTVRLTMQVAVDNAFVVPQAVTRGADDTWMIVNAOGGMEPREVTVAOQGTWIVTSG 357
DB 341 LFLV-----VAGNT-GAPSSIGSPGADSAALTVA-----VDSBRAAVFTSG 383
QY 358 LKDGKVVVEGISINGI 374
DB 384 PRHGDNALKPDLIAPGV 400

RESULT 26
US-09-815-242-5883
Sequence 5883, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5883
LENGTH: 837
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-5883

Query Match 5.6%; Score 114; DB 9; Length 837;
Best Local Similarity 23.3%; Pred. No. 0.2;
Matches 82; Conservative 43; Mismatches 139; Indels 88; Gaps 19;

QY 112 LESARAOATATQATLAKADADLARKP---LVAAEAVSRQSYDAVTAKSAAEG----- 163
DB 18 VEHATTELNTMTLTKALADKATKASGVNVAADAKKQAYDEKVTAAENIVSGTPTPT 77
QY 164 -----VKAQAKISAGINLN-----RSRTAPISGFIG-----OSKXSEGLTNAG 205
DB 78 LTPADVNTAAQTQVNTAKTQNLGNHLEVAKONANTALDGLTSLNGPOKAKKE-----QVG 133
QY 206 DTVLATITRONPMYVAVTQASAEVVK-LRQIAEGKLLAADGVIAVGIKFDDGTVPYE- 263
DB 134 QATLPPVQIVR---DNAQTLLNTAMKGLRDSIAN-----EATIKAGQNTDASQNKQT 183
QY 264 ---KGRLLFADPVNESTG-----QITLRAAVPNDQNIIMPGLVYRVLMQVA 308
DB 184 DYNAAVTAARAKIIIGQTSPSMAQEIINQAKQVTAQOALMGQENL-----RTA 232
QY 309 VDNA-FVVPQAVTRGAKDTWMIVNAOGGMEPREVTVAOQ--QGTWIVTSGKDG--- 361
DB 233 QTNAKQHLNGSLDITLDAQDAVK-ROLEGATHVNEVTOQNNADALNTAMTN-LKNGIOD 290
QY 362 DKVVEGISINGITGAKK-----VTPKEMASSENOAAAP---OSGVOTASE 404
DB 291 QNTIKQGVNFTDADEAKRNAYTNAVTOAE-QIINKAGQFNTSKDGVETALE 340

RESULT 27
US-09-815-242-13080
Sequence 13080, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242

;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; NUMBER OF SEQ ID NOS: 14110
;; SOFTWARE: FASTSEQ for Windows Version 4.0
;; SEQ ID NO: 12996
;; LENGTH: 6281
;; TYPE: PRT
;; ORGANISM: Staphylococcus aureus
US-09-815-242-12996

Query Match 5.6%; Score 114; DB 9; Length 6281;
Best Local Similarity 23.3%; Pred. No. 4;

Matches 82; Conservative 43; Mismatches 139; Indels 88; Gaps 19;

QY 112 LESARAQATATQAATLAKADADLARYK---LVAAEVSRQEYDAVATKRSNAAG----- 163
DB 1399 VEHTATELNTMTALKRAIADAEKTKASGVNNAADANKRQAVDEKVTAAENIVSGTPTPT 1458
QY 164 -----VKAAQAAIKSAGINLN-----RSRTTAPISGFTG-----QSKVSEGLLNAG 205
DB 1459 LTPADVNTAAATQVTAQTQNLGNHNLVAAKQANANTAIIDGLTSLNGPQAKLKE-----QVG 1514
QY 206 DTVYATIRGRTNPMVNTVQSASEVWK-LRROIAGSKLLADGVAVGKFPDGTVYE- 263
DB 1515 QATLTPNVQTVR---DNAQTNTMTAMKGLRDSIAN-----EATTKAQONTYDASQNKOT 1564
QY 264 --KGRLFPADPVNNESTG-----QITRAAVPNDQNLMPGLYVRVLMDOVA 308
DB 1565 DYNNAVTAAKAIIGTGTSPSMAQGINQAKQVTKAQALNGENL-----RTA 1613
QY 309 VDNA--FVVPQQAATVRGADVTVMVNAQSGMEPREVYVAAQ--QGTNMTVTSGLKDG--- 361
DB 1614 QTNNAQHLNGLSDLTDAQDAVK-RQIEGATHVNEVTOQNNAADALNTAMTN-LRNGIOD 1671
QY 362 DKVNVGEISIAIGTGAKK-----VTPKEMASSEQAAAP---QSGVQGTASE 404
DB 1672 QNTIKQGVNFDTDADEAKRNAVYNAVTOAE-QILNAQSGPNTSKQGVETALE 1721

RESULT 30

US-10-156-761-12370
;; Sequence 12370, Application US/10156761
;; Publication No. US20030119018A1
;; GENERAL INFORMATION:
;; APPLICANT: OMURA, SATOSHI
;; APPLICANT: IKEDA, HARUO
;; APPLICANT: ISHIKAWA, JUN
;; APPLICANT: HORIKAWA, HIROSHI
;; APPLICANT: SHIBA, TADAYOSHI
;; APPLICANT: SAKAKI, YOSHIYUKI
;; APPLICANT: HATTORI, MASAHIRA
;; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;; FILE REFERENCE: 249-262
;; CURRENT APPLICATION NUMBER: US/10/156,761
;; CURRENT FILING DATE: 2002-05-29
;; PRIOR APPLICATION NUMBER: JP 2001-204089
;; PRIOR FILING DATE: 2001-05-30
;; PRIOR APPLICATION NUMBER: JP 2001-272697
;; PRIOR FILING DATE: 2001-08-02
;; NUMBER OF SEQ ID NOS: 15109
;; SEQ ID NO: 12370
;; LENGTH: 272
;; TYPE: PRT
;; ORGANISM: Streptomyces avermitilis
US-10-156-761-12370

Query Match 5.6%; Score 112.5; DB 15; Length 272;
Best Local Similarity 24.8%; Pred. No. 0.052;

Matches 72; Conservative 29; Mismatches 112; Indels 77; Gaps 12;
QY 12 AALAAVALVSSCGKGDAAGQGPAGREAPVGVVTHPQTVALTYELPRLSLR 71
DB 24 AGVAGGVSTLVAAGAAG-AANAAEP-----VQTYELP-----TL 58
QY 72 TADVAQVGGIIOKRLFOEGSYVRAGPLVQIDSTVEA-----NLESARAQATATQA 127
DB 59 TADLATQVA-----QSADATQAAANYVELQERDAAAQAAKQA 97
QY 128 KADADLARYKPLVAAEVSRQEYDAVATKRSNAAGKAAQAAIKSAGININRRTAP- 186
DB 98 KSDLDADKKKAETKKAERAAERAAERASASERATLSASAS-ASTGTSTSTSTATYG 156
QY 187 -----ISGFISQSKVSEGLTNAGDTT--VLATIRGRTNPMVNT--QSASEVWKLR 234
DB 157 SAAVIAFVRKQIGDAYVSGGTGPNSDGSLVQAAKQGVSLPRVSQDSTAGTQVSL 216
QY 235 RQIAEGKLL--AADGVAVGIRKFPDGTVPKGRLLFPADVNESTGQI 281
DB 217 SNLQPGDILVYGAGSAVHVGVVGDG-----MFVG-AQNPSTGVV 256

RESULT 31

US-09-815-242-10096
;; Sequence 10096, Application US/09815242
;; Patent No. US20020061569A1
;; GENERAL INFORMATION:
;; APPLICANT: Haselbeck, Robert
;; APPLICANT: Ohlsen, Karl L.
;; APPLICANT: Zyskind, Judith W.
;; APPLICANT: Wall, Daniel
;; APPLICANT: Trawick, John D.
;; APPLICANT: Carr, Grant J.
;; APPLICANT: Yamamoto, Robert T.
;; TITLE OF INVENTION: Identification of Essential Genes in
;; FILE REFERENCE: ELITRA.011A
;; CURRENT APPLICATION NUMBER: US/09/815,242
;; CURRENT FILING DATE: 2001-03-21
;; PRIOR APPLICATION NUMBER: 60/191,078
;; PRIOR FILING DATE: 2000-03-21
;; PRIOR APPLICATION NUMBER: 60/206,848
;; PRIOR FILING DATE: 2000-05-23
;; PRIOR APPLICATION NUMBER: 60/207,727
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/242,578
;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/269,308
;; NUMBER OF SEQ ID NOS: 14110
;; SOFTWARE: FASTSEQ for Windows Version 4.0
;; SEQ ID NO: 10096
;; LENGTH: 405
;; TYPE: PRT
;; ORGANISM: Escherichia coli
US-09-815-242-10096

Query Match 5.6%; Score 112.5; DB 9; Length 405;
Best Local Similarity 22.2%; Pred. No. 0.094;
Matches 90; Conservative 54; Mismatches 142; Indels 119; Gaps 18;

QY 15 AAVALVSSCGKGDAAGQGPAGREAPVGVVTHPQTVALTYELPRLSLRAD 74
DB 14 SVADATVATWTKKGDV-----VRDEVLVEITDKVLEVP----- 50
QY 75 VRAQVGGIIOKRLFOEGSYVRAGPLVQI--DSSTYANLES-----ARAQATATQA- 123

Db 51 --ASADGILDVLEDEBGTIVTSRQILGRLREGNSAGKETSASKEKASTPAQOASIEE 108
Qy 124 -----ATLAKADADLARKYPLVAEAHSVROEYD---AAVTAKRSA-EGGVKAAQ 168
Db 109 QNNDALSPAIRRLAEHNLDAISAIKGTGVGRLTREDVEKHLAKAPAKESAPAAAAPAAQ 168
Qy 169 AAIKAGINLRSRRTTAPISGFIGOSKVSSEGTLLNAGDTTVLATIRQINPMVAVNTQSAS 228
Db 169 PALAA-----RSEKRVPMTRL--RKRVAERLLEAKNSATMLTTNEVY-----MK 211
Qy 229 EVMKLRQIAEGKLLAADGVIAVGIKPDDGTV-----YPEKRLLPADPVNVESTQOIT 282
Db 212 PIMDLRKQYGEA-PEKSHGIRLGFMSFYKAVVEALKRYPEVNASIDGDDVYHNYFDVS 270
Qy 283 LRAAVPNDQNLIMPGLYVRVLMDOVAVDNAFVFPQOAVTRGAKD-----TVMIVNAGGGM 337
Db 271 MAVSTPR-----GLVTPVLRD---VD-----TLGMADIEKKIKELAVKGRDG- 309
Qy 338 EPREVTVAOQOQTNWITVTSGLKDGKRVVVEGISIAGITGAKKVP 382
Db 310 ---KLTVEDLTGNGFTITNG-----GVFGSLMSTP 336

RESULT 32

US-10-078-107-6
; Sequence 6, Application US/10078107
; Publication No. US20020182688A1
; GENERAL INFORMATION:
; APPLICANT: IZUI, HIROSHI
; APPLICANT: HARA, YOSHITAKO
; APPLICANT: SATO, MASAKAZU
; APPLICANT: AKIYOSHI, NAOKI
; TITLE OF INVENTION: METHOD FOR PRODUCING L-GLUTAMIC ACID
; FILE REFERENCE: 219846USO
; CURRENT APPLICATION NUMBER: US/10/078,107
; PRIOR FILING DATE: 2002-05-13
; PRIOR APPLICATION NUMBER: JP 2001-044134
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 405
; TYPE: PRF
; ORGANISM: Escherichia coli
US-10-078-107-6

Query Match 5.6%; Score 112.5; DB 14; Length 405;

Best local Similarity 22.2%; Pred. No. 0.094;
Matches 90; Conservative 54; Mismatches 142; Indels 119; Gaps 18;

Qy 15 AAAVALVLSGCGGDAAGQGPAGREAPVVGVTVPQTVALTVELPGRLESIRAD 74
Db 14 SVADATVATMHHKRGDAV-----VRDEVLVEITDKVLEVP-----50
Qy 75 VRAOVGGIIQKRLFOEGSYRAGOPLYOI---DSSTVEANLES-----ARAQLATAQ- 123
Db 51 --ASADGILDVLEDEBGTIVTSRQILGRLREGNSAGKETSASKEKASTPAQOASIEE 108
Qy 124 -----ATLAKADADLARKYPLVAEAHSVROEYD---AAVTAKRSA-EGGVKAAQ 168
Db 109 QNNDALSPAIRRLAEHNLDAISAIKGTGVGRLTREDVEKHLAKAPAKESAPAAAAPAAQ 168
Qy 169 AAIKAGINLRSRRTTAPISGFIGOSKVSSEGTLLNAGDTTVLATIRQINPMVAVNTQSAS 228
Db 169 PALAA-----RSEKRVPMTRL--RKRVAERLLEAKNSATMLTTNEVY-----MK 211
Qy 229 EVMKLRQIAEGKLLAADGVIAVGIKPDDGTV-----YPEKRLLPADPVNVESTQOIT 282
Db 212 PIMDLRKQYGEA-PEKSHGIRLGFMSFYKAVVEALKRYPEVNASIDGDDVYHNYFDVS 270
Qy 283 LRAAVPNDQNLIMPGLYVRVLMDOVAVDNAFVFPQOAVTRGAKD-----TVMIVNAGGGM 337
Db 271 MAVSTPR-----GLVTPVLRD---VD-----TLGMADIEKKIKELAVKGRDG- 309

Qy 338 EPREVTVAOQOQTNWITVTSGLKDGKRVVVEGISIAGITGAKKVP 382
Db 310 ---KLTVEDLTGNGFTITNG-----GVFGSLMSTP 336

RESULT 33

US-10-077-751-6
; Sequence 6, Application US/10077751
; Publication No. US20020192772A1
; GENERAL INFORMATION:
; APPLICANT: SATO, MASAKAZU
; APPLICANT: AKIYOSHI, NAOKI
; TITLE OF INVENTION: METHOD FOR PRODUCING L-GLUTAMIC ACID
; FILE REFERENCE: 219849USO
; CURRENT APPLICATION NUMBER: US/10/077,751
; PRIOR FILING DATE: 2002-05-13
; PRIOR APPLICATION NUMBER: JP 2001-044135
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 405
; TYPE: PRF
; ORGANISM: Escherichia coli
US-10-077-751-6

Query Match 5.6%; Score 112.5; DB 14; Length 405;

Best local Similarity 22.2%; Pred. No. 0.094;
Matches 90; Conservative 54; Mismatches 142; Indels 119; Gaps 18;

Qy 15 AAAVALVLSGCGGDAAGQGPAGREAPVVGVTVPQTVALTVELPGRLESIRAD 74
Db 14 SVADATVATMHHKRGDAV-----VRDEVLVEITDKVLEVP-----50
Qy 75 VRAOVGGIIQKRLFOEGSYRAGOPLYOI---DSSTVEANLES-----ARAQLATAQ- 123
Db 51 --ASADGILDVLEDEBGTIVTSRQILGRLREGNSAGKETSASKEKASTPAQOASIEE 108
Qy 124 -----ATLAKADADLARKYPLVAEAHSVROEYD---AAVTAKRSA-EGGVKAAQ 168
Db 109 QNNDALSPAIRRLAEHNLDAISAIKGTGVGRLTREDVEKHLAKAPAKESAPAAAAPAAQ 168
Qy 169 AAIKAGINLRSRRTTAPISGFIGOSKVSSEGTLLNAGDTTVLATIRQINPMVAVNTQSAS 228
Db 169 PALAA-----RSEKRVPMTRL--RKRVAERLLEAKNSATMLTTNEVY-----MK 211
Qy 229 EVMKLRQIAEGKLLAADGVIAVGIKPDDGTV-----YPEKRLLPADPVNVESTQOIT 282
Db 212 PIMDLRKQYGEA-PEKSHGIRLGFMSFYKAVVEALKRYPEVNASIDGDDVYHNYFDVS 270
Qy 283 LRAAVPNDQNLIMPGLYVRVLMDOVAVDNAFVFPQOAVTRGAKD-----TVMIVNAGGGM 337
Db 271 MAVSTPR-----GLVTPVLRD---VD-----TLGMADIEKKIKELAVKGRDG- 309
Qy 338 EPREVTVAOQOQTNWITVTSGLKDGKRVVVEGISIAGITGAKKVP 382
Db 310 ---KLTVEDLTGNGFTITNG-----GVFGSLMSTP 336

RESULT 34

US-09-741-669-414
; Sequence 414, Application US/09741669
; Patent No. US2002022718A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; TITLE OF INVENTION: Genes identified as required for
; TITLE OF INVENTION: proliferation of E. coli
; FILE REFERENCE: ELITRA.009A
; CURRENT APPLICATION NUMBER: US/09/741,669
; CURRENT FILING DATE: 2000-12-19

PRIOR APPLICATION NUMBER: US 60/173005
PRIOR FILING DATE: 1999-12-23
NUMBER OF SEQ ID NOS: 481
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 414
LENGTH: 542
TYPE: PRT
ORGANISM: Escherichia coli
US-09-741-669-414

Query Match 5.5%; Score 112; DB 9; Length 542;
Best Local Similarity 24.2%; Pred. No. 0.16; Mismatches 93; Indels 64; Gaps 16;
Matches 65; Conservative 47; Mismatches 93; Indels 64; Gaps 16;

QY 168 QAKISAGINLRSRITAPISGFIGOSKVEGT-----LNAGD-TTVLATIR 214
DB 113 ENGISAERADPOQALIEBAAVNPALIGAVKVPDGVDPRLTAAMMDAKHGAIVLTH 172
QY 215 QTNPMYVNTGASAEV-MKLRRQI-AEGKLLADGVI-AVGI-----KEDDGTV--YPE 263
DB 173 EVTGL---IREGATVGVVRNHLTGCTQALHAPVNVNAGIWMGHIAEYADLRIMFPA 229
QY 264 KGRLLFADPVNYESGQITLRAAVNDONILMPGLV-----RYLMD 305
DB 230 KSLILIMHRINQ---HYINRCRPSDADILVPGDTISLIGTSLRIDYNEIDNRVTAE 286
QY 306 QVAV---DNAFVPPQAVTRGAKDTMTVNAAGMEPREVTVAOQGTMTVTSGLKGD 362
DB 287 EVDILRREKELAPMAKTR-----ILKAYSGRPLVASDDDSGN--VSRGIVLDD 337
QY 363 KVVVEGI-SIAGITAKKVTPK--EMAS 387
DB 338 HAERDGLDGFITITGKLTMTYLAEMWAT 366

RESULT 35

US-09-159-469-50
Sequence 50, Application US/09159469
Patent No. US20020064535A1

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
TITLE OF INVENTION: THERAPY OF EHRlichia INFECTION
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA

ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/159,469
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/106,582
FILING DATE: 29-JUN-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.439C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4800
TELEFAX: 206-682-6031

INFORMATION FOR SEQ ID NO: 50:

SEQUENCE CHARACTERISTICS:
LENGTH: 578 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-159-469-50

Query Match 5.5%; Score 112; DB 9; Length 578;
Best Local Similarity 20.5%; Pred. No. 0.18;
Matches 98; Conservative 64; Mismatches 151; Indels 164; Gaps 18;

QY 6 FKAMRAAA---LAAVALVSSCGKGDAAQGGPAGREAPAPVGVVTHPQTVALTVE 62
DB 134 FKIKSAAEBPQTVDPSSVESATSGVD--TCEOEIOEAPV----- 174
QY 63 LRGLESRTADVAAQVGGIQRKLPFGSVVRAGOPLOYI----- 103
DB 175 ---IEVEETE-----QEVILEGTLIDLEQVPAQVPAVEALPGVEAAEALVP 221
QY 104 ---DSSTVEA-----NLBSARAQLATAQ---ATLAKADADLARVPLVAAEA---VSR 147
DB 222 SLEENKQOEVVVAPEAQLESAPVSAPOPESTVLGVAEGD---KSEVSVEANADVAQ 278
QY 148 QEV-----DAVTAKSAAE-----GVK 165
DB 279 KEVISGOOEQIEAIELEGTEAPVEKETEVLKEDTLIDLEQVPAQVPAVEALPGVE 338
QY 166 AQAIAISAGINLRSRITAPISGFIGOSK-----VSRGTLINADDTVLA 211
DB 339 AAEALVPSLEENKQOEVVVAPEAQLESAPVSAPOPESTVLGVBDDLKSEVSVEADA 398
QY 212 TIROTNPYVNVVNTGASAEV-MKLRRQIABGKLLADGVIAVGIKEDDGTVYPEKGRLLFAD 271
DB 399 GMQOEAGISDQETATEVEVE-----VSVEKTEBEPVILEEGTLIDLE 444
QY 272 -----PVNYES--TGQITLRAAVPNDONILMPGLYRVLMDQVAVDNAFVPPQO----- 318
DB 445 QVPAQVPAVEALPGVEAAEALVPSLEENKQOEVVVAPEAQLESAPVSAPOPESTV 504
QY 319 -AVTRGAKDTMTVNAAGMEPREVTVAOQGTMTVTSGLKGDGVVBEISIAGI 374
DB 505 LGVTGDLKSEVSVEADAGNQ-QEAGISDDE-----TQATEVEKVEVSVEADAGM 554

RESULT 36

US-09-798-042-50
Sequence 50, Application US/09798042
Patent No. US20020068343A1

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
TITLE OF INVENTION: AND TREATMENT OF EHRlichia INFECTION
FILE REFERENCE: 210121.439C7
CURRENT APPLICATION NUMBER: US/09/798,042
CURRENT FILING DATE: 2001-03-02
NUMBER OF SEQ ID NOS: 108
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 50
LENGTH: 578
TYPE: PRT
ORGANISM: Ehrlichia sp.
US-09-798-042-50

Query Match 5.5%; Score 112; DB 9; Length 578;
Best Local Similarity 20.5%; Pred. No. 0.18;
Matches 98; Conservative 64; Mismatches 151; Indels 164; Gaps 18;

QY 6 FKAMRAAA---LAAVALVSSCGKGDAAQGGPAGREAPAPVGVVTHPQTVALTVE 62
DB 134 FKIKSAAEBPQTVDPSSVESATSGVD--TCEOEIOEAPV----- 174
QY 63 LRGLESRTADVAAQVGGIQRKLPFGSVVRAGOPLOYI----- 103
DB 175 ---IEVEETE-----QEVILEGTLIDLEQVPAQVPAVEALPGVEAAEALVP 221
QY 104 ---DSSTVEA-----NLBSARAQLATAQ---ATLAKADADLARVPLVAAEA---VSR 147
DB 222 SLEENKQOEVVVAPEAQLESAPVSAPOPESTVLGVAEGD---KSEVSVEANADVAQ 278
QY 148 QEV-----DAVTAKSAAE-----GVK 165
DB 279 KEVISGOOEQIEAIELEGTEAPVEKETEVLKEDTLIDLEQVPAQVPAVEALPGVE 338
QY 166 AQAIAISAGINLRSRITAPISGFIGOSK-----VSRGTLINADDTVLA 211
DB 339 AAEALVPSLEENKQOEVVVAPEAQLESAPVSAPOPESTVLGVBDDLKSEVSVEADA 398
QY 212 TIROTNPYVNVVNTGASAEV-MKLRRQIABGKLLADGVIAVGIKEDDGTVYPEKGRLLFAD 271
DB 399 GMQOEAGISDQETATEVEVE-----VSVEKTEBEPVILEEGTLIDLE 444
QY 272 -----PVNYES--TGQITLRAAVPNDONILMPGLYRVLMDQVAVDNAFVPPQO----- 318
DB 445 QVPAQVPAVEALPGVEAAEALVPSLEENKQOEVVVAPEAQLESAPVSAPOPESTV 504
QY 319 -AVTRGAKDTMTVNAAGMEPREVTVAOQGTMTVTSGLKGDGVVBEISIAGI 374
DB 505 LGVTGDLKSEVSVEADAGNQ-QEAGISDDE-----TQATEVEKVEVSVEADAGM 554

Db 134 FKLEKSAEFPOTVDSVSESATGSGVD--TQEEQETIDQEPAP----- 174
Qy 63 LFRGLSELTADVRAQVGIIQKRLFOEGSYVRAQGPVQI----- 103
Db 175 ---LEEVTEE-----QEVILEBETLIDLPQVAVVVAEALPGVEAEALVY 221
Qy 104 ---DSSTEA-----NIESAPQALATQ---ATLAKADADLARYKPLVAEA---VSR 147
Db 222 SLEENKQEVVVAPEAQQLSEAPESAPAPQESTVLGVAEGLD---KSEVSEANADVAQ 278
Qy 148 QEV-----DAVTAKRSAEA-----GVK 165
Db 279 KEVISGQOEQETIAELBEGTAPVEVEKETETVLLKEDTLIDLEQPAQVPAVVAEALPGVE 338
Qy 166 AAQAAIKSAGININRSRITAPISGFIGSK-----VSEGLTNAGDTYLA 211
Db 339 AAELVPLSEENKQOEVVVAPEAQQLSEAPESAPAPQESTVLGVTEGDLKSEVSEVADA 398
Qy 212 TIRQTNPMYVAVTQASAEVVKLRQIAGKLLAADGVIAVGIKFDDGVYPERKRLFPAD 271
Db 399 GMOQFAGISDQETQATEVEKE-----VSEVETKTEPEVILEBETLIDLE 444
Qy 272 -----PVVNES--TGQTLTAAVNDQNIIMPLGLYVAVLMDQVAVNDVAFVVPQ----- 318
Db 445 QPVAQVPAVVAEALPGVEAEALVPSLEENKQOEVVVAPEAQQLSEAPESAPAPQESTV 504
Qy 319 -AVTRGADTVMIVNAQGMREPVTVAAQOQGTNMTVSGLKDGDKVVEGISIAGI 374
Db 505 LGVTEGDLKSEVSEVADAGM-QEAGISDQ-----TQATEVEVAVKVSVEADAGM 554

RESULT 37

US-09-815-242-5885
; Sequence 5885, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zykend, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5885
; LENGTH: 1029
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5885

Query Match 5.5%; Score 111; DB 9; Length 1029;
Best Local Similarity 21.2%; Pred. No. 0.51;

Matches 84; Conservative 58; Mismatches 130; Indels 124; Gaps 19;
Qy 58 ALTVELPERLESLRTADVRAQVGIIQ---KRLFOEGSYVRAQGPVQIDSTYEANE 113
Db 35 ARAIDALPHLNNAQADVAKSKINAAASNAGVNTVQOQT-----DLNTMGNLQ 83
Qy 114 SARAQALATQATLAKADADLARYKPLVAEAVSROEYDAATAKRSAAE-GVKAQAAI- 171
Db 84 GA---INDEQTLT-----NSQNVQDAPSPSKTATVNAVQAADKLTN 121
Qy 172 KSAGININRSRITAPISGFIGSKVSEGLT-----NAGDTYLATIRQTNPMYV 221
Db 122 KSNQNKTKDQVTEAMQVNSAKXNLDGTRLLDQAKQAKOQLNMNTHLTJAQKTN---- 177
Qy 222 NVTQASAVMKLRQIAGKLLAADGVIAVGIKFDDGVYPERKRLFPADVMESTQI 281
Db 178 -----LTQINSQTTVA-----GVQ-----TVQSNANTL---DQAMN----- 206
Qy 282 TLRAAVPNDQNIIMPLGLYVAVLMD-QVAVNDVAFVVPQAVTRGADTVMIVNAQGMREP- 339
Db 207 TLRSQANKKATKASQEDVYDANNKQATYNNAV-----AAETIINNSNPEMPS 257
Qy 340 -----REVTVAAQOQGTNMTVYV-SGLKDGDKVVEGISIAGITGAKK 379
Db 258 TITQKAEQVNSKTLANGDENILTPAKQAKTYLNTLISITDAK---NNLISQITSATR 313
Qy 380 V-----TPKEWASSENQA-AAPOSGVQTASEAKTASE 410
Db 314 VSGVDTVKQNAQHLQAMASLQNGINNESQVKSSEK 349

RESULT 38

US-09-815-242-13083
; Sequence 13083, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zykend, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13083
; LENGTH: 1048
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-13083

Query Match 5.5%; Score 111; DB 9; Length 1048;
Best Local Similarity 21.2%; Pred. No. 0.53;

Matches 84; Conservative 58; Mismatches 130; Indels 124; Gaps 19;

QY 58 ALTYELPERLSIRADYRAOVGIIQ-----KRLFOEGSYRAGQPLVQIDISTYEANLE 113
 DB 47 ARAIDALPHLNNAQADVSKSKINAASNTAGVNTVQOQST-----DLNTMGMLQ 95
 QY 114 SARAQATATQATLAKADADLARKEPLVAEAASROEYDAAVTAKSAAE-GVKAQAQAI- 171
 DB 96 GA-----INEQRTL-----NSQNYQDAIPSKSKTATVNAVQAADILN 133
 QY 172 KSAGINLNRSRITAPISGFIGOSKVSSEGLT-----NAGCTVTLATIRQTNPMYV 221
 DB 134 KSMGNKTKDQVTEAMQVNSAKXNNLDGIRLIDQAKQAKQOLNMTHLITTAOKIN----- 189
 QY 222 NVTQASAVMKLRQIAGSKLLAADGVIAVGKFPDGTVPYEPKRLLPADPVNESTQOI 281
 DB 190 -----LTIQINSGITVA-----GVQ-----TVQSNANTL---DQAMN----- 218
 QY 282 TLRAAVPNDONILMGLYVRVLM-DQAVDNAFVVPQOAVTRGAKDTWIVNAOGGMP- 339
 DB 219 TLROSIANKDATKASQEDYVDANNQKQTAYNNNAV-----AAEITLIINNSNPEMPS 269
 QY 340 -----REYTVAAQOQGTWIVT-SGLKDGKVVVEGISTAGITGAKK 379
 DB 270 TITQKAEQVNSKTLANGDENILTAQONAKTYLANTLITSIDAQK-----NNLISQITSATR 325
 QY 380 V-----TPKEMASSENQA-AAPOSGVQTASEAKTASE 410
 DB 326 VSGVDTVRKNAQHLDOAMASLQNGINNESOVKSSK 361

RESULT 39
 US-09-782-714-1
 ; Sequence 1, Application US/09782714
 ; Patent No. US20020034808A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Dazet, Seth A
 ; APPLICANT: Zhang, Gongyi
 ; APPLICANT: Campbell, Elizabeth
 ; APPLICANT: Minikin, Leonid
 ; APPLICANT: Severinov, Konstantin
 ; TITLE OF INVENTION: A CRYSTAL OF BACTERIAL CORE RNA POLYMERASE AND METHODS
 ; FILE REFERENCE: 600-1-258
 ; CURRENT FILING DATE: 2001-02-13
 ; PRIOR APPLICATION NUMBER: US/09/782,714
 ; PRIOR FILING DATE: 1999-09-15
 ; NUMBER OF SEQ ID NOS: 4
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 1525
 ; TYPE: PRT
 ; ORGANISM: Thermus aquaticus
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (1247)
 ; OTHER INFORMATION: Any amino acid can be at this position
 US-09-782-714-1

Query Match 5.5%; Score 111; DB 9; Length 1525;
 Best Local Similarity 19.8%; Pred. No. 0.92;
 Matches 98; Conservative 73; Mismatches 161; Indels 164; Gaps 25;

QY 21 VLSGCKGDAAGQ-----PAGREAPAV-----GVVTVHPQVALTVLPG 65
 DB 1023 VMAQSGARGNPQIRQLCGMRGLMQPSEGFPEVVRSSFRGLTVLEVFISSHGARKG 1082
 QY 66 RLSEKRTADVRAQVGGIIQKRLFOE-----GSVRAQGPLYQIDISTYANLES 114
 DB 1083 ADTALRTID-----SGYLTRKLVDAHEIVREADCGTTKYISVLPFQWDEVTRTLRLRK 1137
 QY 115 ARAQATATQATLAKADADLARKEPLVA--AEAVSROEYDAAVTAKSAAEAGVKAQA--- 169

DB 1138 -----RSDIESGLYGRVLAREVEALGRRLIEGRYSLLEDVHFLITKAAGEV 1184
 QY 170 -----AISKAGINLNRSRITAPIS-----GFIGOSKVSF-GTLL----- 202
 DB 1185 REVERSPPLTCQTRYGVCCQKYGIDLSMAR---PVSIGEAAGVAABSIGEPQTLMRT 1241
 QY 203 -NAGDTVTLATIRQTNPMYVNVVTQASAEVVKLRQIAGSKLLAADGVIAVGKFPDGTVPY 261
 DB 1242 FHTGXAAGVDITDGLRVI-----ELFEARRPKAKAVISEIDGV---VRIEFG--- 1287
 QY 262 PEKGRLLPADPVNESTQOITLRAAVPNDONILMGLYVRVLMQAVDNAFVVPQOAVT 321
 DB 1288 --EDRL-----SVFESEEG-FSKYEKLPKDAARLVK-----DGDYVEAGQPLT 1327
 QY 322 RGAKDTWIVNAAG-----GMEPREVTVAAOQGTWVI-VT----- 355
 DB 1328 RGAIDPHQLEAKKPEAVERLYVDEIQKVRAGGVKLDKHIIEIVRQMLKRYEVTDPGD 1387
 QY 356 SGLKDG-----DKVVVEG-ISA-----GITGAKKVPKEMASSENQAAPQ 396
 DB 1388 SPLLEGQVLEKMDVEALNERLIAEGKVPVAMKPLMGVT-KSALSTKSWLS---AASFQ 1442
 QY 397 SGVQTASEAKTASEAE 412
 DB 1443 NTHVLTAAIALAGKGD 1458

RESULT 40
 US-10-156-761-7965
 ; Sequence 7965, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; CURRENT FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 7965
 ; LENGTH: 7746
 ; TYPE: PRT
 ; ORGANISM: Streptomyces avermitilis
 US-10-156-761-7965

Query Match 5.5%; Score 111; DB 15; Length 7746;
 Best Local Similarity 20.9%; Pred. No. 10;
 Matches 97; Conservative 65; Mismatches 162; Indels 140; Gaps 20;

QY 41 EAPAPVVGVT-----VHPQVAL-----TVLPGRLSIRTA 73
 DB 2901 QAAAGVAGVIGVWLAMRHGVLPQTLHVEDPTAHVMSAGAVRLLTPEVQWGTGPRRAA 2960
 QY 74 DYRAQVGGIIQKRLFOESYRAGQPLVQ-----IDSTYANLES-ARAQATATQA 124
 DB 2961 VSSFQIGGTNAHTIIEEPAAAAAEPAAREHNPVVPWVLSKSSVALRAQGERLLAFTD 3020
 QY 125 TLKADADLARKEPLVAEAASROEYDAAVTAKSAAE--AGVKAQAQAIKSGINLNRSR 182
 DB 3021 DVSEVDAQFS-----SATTSALFHRAAVIGSPAEIRAGLELMAAGEPAANVYAGRAH 3074
 QY 183 ITAPISGFI-----GOSKVSSEGLTLNAG-----DTTV---LATIRQTNPMY 220


```

Db      3075 -SADKVGFLFSGQGSQRLGMGBELYAAYPFPAAYDEVCALLDATVDVDAETLHQTGS-- 3131
QY      221  VNVTSASEVMKLRROIAGKLLADGV-----IAGIKFDGDTYPE 263
Db      3132 ---TQPALFAV-----EVALFRLLESWGIRPDYVAGHSVGEIAAAHVAGVLSLEDA----- 3179
QY      264  KGRLLFADPVNNESTGQITLRAAVPNDQNIIMPGLYVRVMDQYAVDNAFVV---PQQA 319
Db      3180 -AKLVSAARAALMKALPAGGVWVAVOAAEDEMPLPHL-----IDRAGIAIANGPRSV 3228
QY      320  VTRGAKDTVMIVNAGGMEPREVVAQOQGTNMIVTS-----GLKDGKVVVEGIS 370
Db      3229 VTSGEDAVTAI-----AEVFAAGRKTSRLKVSASHFSPMLMDPMLLEEFVAVRGLT 3280
QY      371  -----IAGITG--AKVTPKEWASSENOAAAPOSQVOTASE 404
Db      3281 FSEPMIPVSNLTGRLTESYTPPEYVVRHVRREAVRFADGVLTLHE 3324

```

Search completed: September 8, 2003, 14:07:25
 Job time : 29 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2003, 13:59:51 ; Search time 21 Seconds
(without alignments)
830.099 Million cell updates/sec

Title: US-09-889-756a-2
Perfect score: 2019
Sequence: 1 MAFYAFKRAAALAAVAAL.....AAPOSGVOTASEAKTASEAE 412

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, AA.*

1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PTCUS.COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfltest.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	802.5	39.7	424	US-09-328-352-4199	Sequence 4199, Ap
2	753	37.3	453	US-09-252-991A-16988	Sequence 16988, A
3	698.5	34.6	410	US-09-252-991A-30590	Sequence 30590, A
4	647	32.0	447	US-09-328-352-7383	Sequence 7383, Ap
5	642	31.8	414	US-09-252-991A-27975	Sequence 27975, A
6	451.5	22.4	409	US-09-328-352-4570	Sequence 4570, Ap
7	448	22.2	455	US-09-252-991A-31636	Sequence 31636, A
8	429	21.2	429	US-09-252-991A-31241	Sequence 31241, A
9	393	19.5	502	US-09-252-991A-24388	Sequence 24388, A
10	370.5	18.4	458	US-09-199-637A-149	Sequence 149, App
11	355	17.6	458	US-09-252-991A-17691	Sequence 17691, A
12	277	13.7	394	US-09-252-991A-24485	Sequence 24485, A
13	272.5	13.5	454	US-09-328-352-7897	Sequence 7897, Ap
14	267	13.2	399	US-09-252-991A-20610	Sequence 20610, A
15	256	12.7	404	US-09-252-991A-28543	Sequence 28543, A
16	245.5	12.2	389	US-09-252-991A-21202	Sequence 21202, A
17	237.5	11.8	770	US-09-252-991A-28510	Sequence 28510, A
18	220	10.9	500	US-09-328-352-8183	Sequence 8183, Ap
19	216.5	10.7	367	US-09-328-352-5003	Sequence 5003, Ap
20	215	10.6	487	US-09-252-991A-25447	Sequence 25447, A
21	197.5	9.8	306	US-09-252-991A-19591	Sequence 19591, A
22	184.5	9.1	367	US-09-328-352-7523	Sequence 7523, Ap
23	184	9.1	423	US-09-328-352-6273	Sequence 6273, Ap
24	184	9.1	517	US-09-252-991A-31115	Sequence 31115, A
25	183.5	9.1	354	US-09-252-991A-18501	Sequence 18501, A
26	179	8.9	356	US-09-328-352-7671	Sequence 7671, Ap
27	174.5	8.6	358	US-09-328-352-5910	Sequence 5910, Ap

28	174.5	8.6	510	US-09-252-991A-25238	Sequence 25238, A
29	173.5	8.6	495	US-09-252-991A-23319	Sequence 23319, A
30	173.5	8.6	365	US-09-328-352-7412	Sequence 7412, Ap
31	171	8.5	386	US-09-328-352-4722	Sequence 4722, Ap
32	168	8.3	336	US-09-328-352-8202	Sequence 8202, Ap
33	164.5	8.1	314	US-09-252-991A-20336	Sequence 20336, A
34	161.5	8.0	357	US-09-252-991A-17970	Sequence 17970, A
35	161.5	8.0	399	US-09-252-991A-23589	Sequence 23589, A
36	161	8.0	373	US-09-252-991A-28902	Sequence 28902, A
37	157.5	7.8	374	US-09-328-352-5275	Sequence 5275, Ap
38	155.5	7.7	449	US-09-252-991A-28283	Sequence 28283, A
39	153	7.6	397	US-09-252-991A-23027	Sequence 23027, A
40	151.5	7.5	377	US-09-328-352-7154	Sequence 7154, Ap
41	148.5	7.4	420	US-09-252-991A-30038	Sequence 30038, A
42	148	7.3	411	US-09-328-352-6719	Sequence 6719, Ap
43	143	7.1	379	US-09-252-991A-26160	Sequence 26160, A
44	138.5	6.9	416	US-09-252-991A-19218	Sequence 19218, A
45	137	6.8	417	US-09-252-991A-28413	Sequence 28413, A

ALIGNMENTS

RESULT 1
US-09-328-352-4199
Sequence 4199, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Bretton et al.
TITLE OF INVENTION: NOCUTIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328.352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 4199
LENGTH: 424
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-4199

Query Match	39.7%	Score 802.5;	DB 4;	Length 424;
Best Local Similarity	43.7%	Pred. No. 2.2e-70;		
Matches 181;	Conservative 80;	Mismatches 136;	Indels 17;	Gaps 8;
QY	8	AMRAALAAVALVSSCGKGDAAOGGPRGRAPVGVVTHVHTVATLTTELRL	67	
DB	18	ALTRCALATSTIALV--GCSKSGDDEKQAAAAQKMPAE-VGVIVAQPQSVQSVLSGRT	74	
QY	68	ESLRTADVRAQVGGIIQKRLFOEGSYVRAGOPLYQIDSYTYEAMLESARQATATQAATIA	127	
DB	75	SAYISEVRQTSVILLKRLFAESYVRREGALYELDSRTIRATLENAKSLLOOQANLA	134	
QY	128	KADADLARVPLVAEAVSRQEDVAATKRSAGVKAQAQAIKSGINLRSRTAPI	187	
DB	135	SLRFLKRLRYQVLVSNAVSKQEVYDILGGVNVAAQVAAKAQVTNNVLDGYSTISPI	194	
QY	188	SGRFGQKRVSEGLTANAGDTTVALTIROTNPMTYVNTQSAEVMKLRQIIEGULLADG	247	
DB	195	SGQSGRSVAGALVTNNQDPLVTIOQLPIYVDINQSSLELRLRQOOLSKGSLNNSNN	254	
QY	248	VIAVGIFEDQGVYPEKGRLLFADPVVNESTGQITLRAAVPNDONIIIMPGLYVVLMDQV	307	
DB	255	T-KYKALIBDSYPIFGQLAFSPASVNDGTITTLAIVSNPHMLLPKMYTTAOLVQGS	313	
QY	308	AVDNAFVFPQOAVTR--GAKDTWIVNAOGGMEBEVTVAAQOQSTNMIYISGLKDGKVV	365	
DB	314	VVPNAVYIIPQALIRLPTGQAVMLVNAKGVESRPVETSGVGOQNMIVTNGLAADGKVI	373	
QY	366	VEGISTIGTIGAKVTPKEMAS--SENQAAAP-----QSGVOTASEAKTASEA	411	
DB	374	VVG--VAKVKEGQVSAKPYQAPANSQGAAPNAKPAQSG-KPOABOKAASNA	424	

```
RESULT 2
US-09-252-991A-16988
; Sequence 16988, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16988
; LENGTH: 453
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16988

Query Match
Best Local Similarity 37.3%; Score 753; DB 4; Length 453;
Matches 174; Conservative 68; Mismatches 125; Indels 46; Gaps 10;

QY 8 AMRAAALAAVAL-VLSSCGKGDAAGGAPGAPAPV---VGVTVHPQTALVYE 62
DB 76 AMRVLPVLPVLAVALISLSCGKGS-----EAPPAQTPPEVGIIVLEAQTVTLNTE 123
QY 63 LFGRLSELTADVRAVQVGGIIQKRLFOEGSYVRAGQPLVQIDSTYEANLESARQALATA 122
DB 124 LFGRLNARFIAVRRQVNGIILKRLFKESDVKAGQQLYQIDPATYADYQAGQANLAST 183
QY 123 QATLAKADADLARYKPLVAAEAVSRQEDAAVTAKRSAGVKAQAALIKSAGINLMSR 182
DB 184 Q-----EQAQRVYKLVADQAVSKQY-----ADANAVALQSKAAVEQARIMLRATK 229
QY 183 ITAPISGFIGSKVSEGTLLNAGDTTVLATIRQTNPMYVNTQASSEVWKLROIAEGTL 242
DB 220 VLSPISGRIGRSATVEGALVTNGQANAMATVQDLPIYDVTPSTALLRLRELASQGL 289
QY 243 -LAAQGVIAVGKIFPDGTVYPERGRLLFADPVVNESTGQITLRAVPNDONILMGLYVR 301
DB 290 EAGAGMAAKVSLKLEDGQYPLEGRLEPSEVDEBGSVTRAVFPNPNMELLFGMPFH 349
QY 302 VLMDQVAVDNAAFVPPQAVTRGAKD--TVMIYNAQSGMEPREVTVAQQQGTWIVTSGLK 359
DB 350 AQLQGVKQKALILAPQGVTRDLKGOATLAVNNAQKVELRVIKADRVIGDKMLVTEGLN 409
QY 360 DGDKVVVVGISIA--GITCAKAVTYKEMASSENQAAPOSGVQTSSEATASE 410
DB 410 AGDKIITTEGLQFVQPGVE-VKTVPAKNVASQAKADAAAP-----AKTDSK 452
```

```
US-09-252-991A-30590
; Sequence 30590, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30590

RESULT 3
US-09-252-991A-30590
; Sequence 30590, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30590

Query Match
Best Local Similarity 42.1%; Pred. No. 1.8e-65;
Matches 174; Conservative 68; Mismatches 125; Indels 46; Gaps 10;

QY 8 AMRAAALAAVAL-VLSSCGKGDAAGGAPGAPAPV---VGVTVHPQTALVYE 62
DB 76 AMRVLPVLPVLAVALISLSCGKGS-----EAPPAQTPPEVGIIVLEAQTVTLNTE 123
QY 63 LFGRLSELTADVRAVQVGGIIQKRLFOEGSYVRAGQPLVQIDSTYEANLESARQALATA 122
DB 124 LFGRLNARFIAVRRQVNGIILKRLFKESDVKAGQQLYQIDPATYADYQAGQANLAST 183
QY 123 QATLAKADADLARYKPLVAAEAVSRQEDAAVTAKRSAGVKAQAALIKSAGINLMSR 182
DB 184 Q-----EQAQRVYKLVADQAVSKQY-----ADANAVALQSKAAVEQARIMLRATK 229
QY 183 ITAPISGFIGSKVSEGTLLNAGDTTVLATIRQTNPMYVNTQASSEVWKLROIAEGTL 242
DB 220 VLSPISGRIGRSATVEGALVTNGQANAMATVQDLPIYDVTPSTALLRLRELASQGL 289
QY 243 -LAAQGVIAVGKIFPDGTVYPERGRLLFADPVVNESTGQITLRAVPNDONILMGLYVR 301
DB 290 EAGAGMAAKVSLKLEDGQYPLEGRLEPSEVDEBGSVTRAVFPNPNMELLFGMPFH 349
QY 302 VLMDQVAVDNAAFVPPQAVTRGAKD--TVMIYNAQSGMEPREVTVAQQQGTWIVTSGLK 359
DB 350 AQLQGVKQKALILAPQGVTRDLKGOATLAVNNAQKVELRVIKADRVIGDKMLVTEGLN 409
QY 360 DGDKVVVVGISIA--GITCAKAVTYKEMASSENQAAPOSGVQTSSEATASE 410
DB 410 AGDKIITTEGLQFVQPGVE-VKTVPAKNVASQAKADAAAP-----AKTDSK 452

RESULT 4
US-09-328-352-7383
; Sequence 7383, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; PRIOR FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7383
; LENGTH: 447
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7383

Query Match
Best Local Similarity 32.0%; Score 647; DB 4; Length 447;
Matches 146; Conservative 66; Mismatches 136; Indels 10; Gaps 3;

QY 14 LAAAVATVLSGCGKGDAAGGAPGAPAPVGVTVHPQTALVTELPERLESRTA 73
DB 60 LFLSLGILLQGC-----DSKEVQAQEPAPKAVSLISQSVNFSENLPARVAFRTA 112
QY 74 DVRAQVGGIIQKRLFOEGSYVRAGQPLVQIDSTYEANLESARQALATQATLAKADL 133
DB 113 EIRPQVGGIIKRVLFKQSGSEVRAGQALYKINSETFEADVNSRSLNKAEEVRLKYOL 172
QY 134 ARYKPLVAAEAVSRQEDAAVTAKRSAGVKAQAALIKSAGINLMSRRTAPISG 193
DB 173 ERYEQLLPSNAISKQEVNNAQOYRQALADVAQKALLARQNLQVYTVAPISGRIGQ 232
QY 194 SKVSEGTLLNAGDTTVLATIRQTNPMYVNTQASSEVWKLROIAEGKLLAADGVIAVG 253
```

Db 233 SFYTEGALVGGQDNTNMATIIQQIDKAYVVDKVSSEYERLQAAQSGE--LSANDKTVRI 291
 QY 254 KPDDGYIYEPKGRLLFADPVYNVESTGQITLLRAVPNDQNIIMPGLYRVLMDOYAVDNAF 313
 Db 292 TNSHGQPYNTAKMLFEDINVDPETDVFYRIEVNNTSKLLPGYRVYVINDRASIPQAL 351
 QY 314 VVPGQAATR--GAKDVTMTVNAQGMREPREVVAQQGQNMNIVTSGLKDGKGVVBEI 369
 Db 352 LVPRQAQQRNITSGEPQYVYVNAQGTALNPIPIITIGQYQGFYANKGLGVGVKRVVEI 409

RESULT 5
US-09-252-991A-27975

```

: GENERAL INFORMATION:
: APPLICANT: Marc J. Rubenfield et al.
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
: TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 107196.136
: CURRENT APPLICATION NUMBER: US/09/252,991A
: PRIORITY FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIOR FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 60/094,190
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ ID NOS: 33142
: SEQ ID NO 27975
: LENGTH: 414
: TYPE: PR1
: ORGANISM: Pseudomonas aeruginosa
: US-09-252-991A-27975

```

Query Match	31.8%;	Score 642;	DB 4;	Length 414;	.
Best Local Similarity	39.2%;	Pred. No. 1.2e-54;			
Matches 157;	Conservative 65;	Mismatches 157;	Indels 22;	Gaps 5;	

RESULT 6
US-09-328-352-4570
; Sequence 4570, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.

```

/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
/ TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: GTC99-03PA
/ CURRENT APPLICATION NUMBER: US/09/328,352
/ CURRENT FILING DATE: 1999-06-04
/ NUMBER OF SEQ ID NOS: 8252
/ SEQ ID NO 4570
/ LENGTH: 409
/ TYPE: PRT
/ ORGANISM: Acinetobacter baumannii
US-09-328-352-4570

```

Query Match	22.4%;	Score 451.5;	DB 4;	Length 409;
Best Local Similarity	31.6%;	Pred. No. 6.3e-36;		
Matches 131; Conservative	82;	Mismatches 180;	Indels 21;	Gaps 7;

```

RESULT 7
US-09-252-991A-31636
: Sequence 31636, Application US/09252991A
: Patent No. 6551795
: GENERAL INFORMATION:
: APPLICANT: Marc J. Rubenfield et al.
: TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO PSEUDOMONAS
: TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 107196.136
: CURRENT APPLICATION NUMBER: US/09/252.991A
: PRIORITY FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIORITY FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 60/094,190
: PRIORITY FILING DATE: 1998-07-27
: NUMBER OF SEQ ID NOS: 33142
: SEQ ID NO 31636
: LENGTH: 455
: TYPE: PR1
: ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31636

```

Query Match	22.2%;	Score 448;	DB 4;	Length 455;
Best Local Similarity	32.7%;	Pred. No. 1.7e-35;		
Matches 133;	Conservative 75;	Mismatches 177;	Indels 22;	Gaps 9;

14 LAAAVLVSSCGGGAACGGAGAEAPVGVTVTHQYVALVELPGRLES LTA 73
 ||| ||||| : ||| ||| ||| : ||| ||| :

```

Db      58 LALAALVLSACGKAPETTG-----MAAPKSVAEVIEQPLNEMDEFTGRLEAPSPV 110
Qy      74 DYRAVGGIGICKRLFOEGSVYRAGPLVYOIDSSTYEANLESARQALATQAATLADADL 133
Db      111 ELRPVSGTIDVAHREGALVKKGDLFQIDPRPEAEVKRLEAOVQQAQAQASVWEA 170
Qy      134 ARYKPLVAEAVSRQEDYDAVTAARSAEAGVAAQAATKAGININRSRITAPISGFIQ 193
Db      171 QRGELRLASMAISAFADARTTAQBAKAAVAATQAQDAALNLSFTRITPIDGRYSR 230
Qy      194 SKVSEGTLLMAGDTTLATIRQTNPMYVNTQASAEVWK--LRQIAEGKLLAADGVIA 250
Db      221 AEVTAGNIVNSGE-TLITLVSTDKVYAFDADERVFLKYVLAHQ--AGRDTRESEPV 287
Qy      251 VGIRKDDGVYPERKRLLEADPVNESGQITLRAAVPDQIILMPGLYVRVLMQVAVD 310
Db      288 LGLSESDGN--PHLGRLDLDVQVNPRTGTRGRAVFDKAKEFPGLNVLKLGSKTY 345
Qy      311 NAFVVPQOAV--TRGAKDTVMIVNAAGMEPREVTVAQOQGTNMIYTSGLKDGKTYVEGI 369
Db      346 AATILKDEAVGTDGKKFVLVDGDKNTYRTVEKPGKLEGRIRYRSGLSKGDRIYVAGL 405
Qy      370 SIAGITGAKKVTPK--EWAASSENQA--AAPQSGVOTASEAKTASAE 412
Db      406 Q--RVRPGMVDPPQKVEWASADTLATLARLRQSVGDSDEPPKVAASKD 450

```

RESULT 8
US-09-252-991A-31241
; Sequence 31241, Application US/09252991A
; Patent No. 6551795

```

; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31241
; LENGTH: 429
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-31241

```

Query Match 21.2%; Score 429; DB 4; Length 429;
Best Local Similarity 31.0%; Pred. No. 1,1e-33;
Matches 119; Conservative 59; Mismatches 172; Indels 34; Gaps 7;

```

Qy      6 FKAMR-----AAALAAVALVL-----SSCGKGDPAQGGOPAGR--EAPAPVGVVTV 52
Db      13 FRTLAPMLITALFAAIVGLVWMLAAPASPSDDPRGKGRGALPRAALITGVGARV 72
Qy      53 HPQVALVELPGRLESLEFADVRAQVGGITIOKRLFOEGSVYRAGPLVYOIDSSTYEANL 112
Db      73 EGGDLALHFNALGTAFVTAFTVAVKPRVNGELVKVLFQEQGEVAKAGILLAVVDPRTYKAL 132
Qy      113 ESARQALATQAQTLTAKADADLARAYKPIVAEAVSRQEDYDAVTAARSAEAGKAAQAALIK 172
Db      133 AQAEGTLLMOMQAOLKNAETIDLRKGLYAEADSIKQTLDTQEAQVRLQGLRTTQGGVD 192
Qy      173 SAGININRSRITAPISGFIQSGKSEGTLLMAGDTTLATIRQTNPMYVNTQASAEVWK 232
Db      193 DARLALITFEVRAPIISGRIGLRQVNDIGNLVTSQDTTLPLVITQVPEISVFSLEPQQIGT 252
Qy      233 LRRQI--AEKLLAA-----DGVIAVGKFDGDTYPERKRLLEADPVNESGQITTLRA 285
Db      253 VVEQNNNGRGLAVTLADRNQDKVLA-----EGTLTLTDNQIDTDTTGVKVLKA 299

```

```

Qy      286 AVPDNQINIMGLYVRVLMQVAVDNAFVVPQAVTRGAKDT--VMIVNAAGMEPREVTV 344
Db      300 RFENADGKLFNQVNVNRLAHLTKGVLTIPANAVQGTNGIYVVVGADKVKQSRSVAI 359
Qy      345 AQOQGTNMIYTSGLKDGKTYVEG 368
Db      360 GTSENERVYVESGLKAGEQVAVVEG 383

```

RESULT 9
US-09-252-991A-24388
; Sequence 24388, Application US/09252991A
; Patent No. 6551795

```

; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24388
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-24388

```

Query Match 19.5%; Score 393; DB 4; Length 502;
Best Local Similarity 33.7%; Pred. No. 4.9e-30;
Matches 118; Conservative 50; Mismatches 162; Indels 20; Gaps 5;

```

Qy      28 GDAAGGQGPARGAPAPV--VGVTVHPQTVALTPEGRLESLEFADVRAQVGGITIQ 84
Db      141 GGMAMLGDDAGAKAPAPARVPVAVRVERDVEQVSGIGTVSLHNVVIRTQIDGQLT 200
Qy      85 KRLFOEGSVYRAGPLVYOIDSSTYEANLESARQALATQAATLADADLARAYKPLVAEA 144
Db      201 RLIVSEGGQVBAEGLATIDRAVVAALQQAASRASQAOLKSAEODLQHRSLYARA 260
Qy      145 VSRQEDYDAVTAARSAEAGVAAQAATKAGININRSRITAPISGFIQSGKSEGTLLNA 204
Db      261 VSRQLDDQQAATVDQLRATTLKANDATTIAEVRRLSYTRITSPVSGKVGIRAVDGNLVRV 320
Qy      205 GDTTVLATIRQTNPMYVNTQASAEVWKLRQI--AEKLLAADGVIAVGIRKFDGTVY 261
Db      321 GDSIGLFEVTLQAPISVFSIQEQQLPQLQALGGAAVRAVSRDGSALG----- 371
Qy      262 PEKRLLEADPVNESGQITLRAAVPDQIILMPGLYVRVLMQVAVDNAFVVPQAVT 321
Db      372 --EGRLTLTDNQIDSSTQITIRVRASFDRQARLMPGQFVAVSLHTGVARDLVSSKXVR 429
Qy      322 RGAKDTVMIVNAAGMEPREVTVAQO--QGTNMIYTSGLKDGKTYVEGIS 370
Db      430 RGLSENFYRYAADREAVPVRVLQDIDGLS--VVEGLASGDQVVDGHS 477

```

RESULT 10
US-09-199-637A-149

```

; Sequence 149, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick
; APPLICANT: Goodman, Howard M.
; APPLICANT: Rahme, Lawrence G.
; APPLICANT: Mahajan-Miklos, Shalina
; APPLICANT: Tan, Man-Wah
; APPLICANT: Cao, Hui
; APPLICANT: Drenkard, Eliana
; APPLICANT: Tsongalis, John

```

```

; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; FILE REFERENCE: SEQUENCES AND USES THEREOF
; CURRENT APPLICATION NUMBER: 00786/361002
; CURRENT FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 149
; LENGTH: 202
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-149

```

```

Query Match      18.4%; Score 370.5; DB 4; Length 202;
Best Local Similarity 37.9%; Pred. No. 1.9e-28;
Matches 80; Conservative 43; Mismatches 73; Indels 15; Gaps 5;

```

```

QY 205 GDTVLATIRQTNPMYVAVTOSASEVMKLRQIAEGKL-LADGVIAVGIRKPDGTVPY 263
DB 1 GOANMATVQQLDPIYVDTQSTALIRMRRELASGQERAGDNAAKVSLEKLEDSQYPL 60
QY 264 KGRLLFADPVYNESGQITLRAAVNDQNIIMPGLYRVLDQVAVDNAFVVPQAVTRG 323
DB 61 EGRLEFSEVDESGVTIRAFVFPNNNELPGFVHAQLOEGVKOKAIIAPQOGVTRD 120
QY 324 AKD--TWIIVNAQGMREPREVVAQOGGTNMTVTSGLKDGKVVVEGSI--GITGAKK 379
DB 121 LKGAQATLVANQKQVRELIVKADRVIGDKMLVTGLNAGDKITTEGLOFVQGVVE-VKT 179
QY 380 VTPKEMASSENQAPAPQSGVQTAASEAKTASE 410
DB 180 VPAKVVASAQKADAP-----AKTDSK 201

```

```

RESULT 11
US-09-252-991A-17691
; Sequence 17691, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196,136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17691
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17691

```

```

Query Match      17.6%; Score 355; DB 4; Length 458;
Best Local Similarity 27.6%; Pred. No. 2.3e-26;
Matches 102; Conservative 78; Mismatches 169; Indels 20; Gaps 7;

```

```

QY 18 VALVSSCGKGDDAAGGQAPGRAPVAVGVVTHPQTVALTVELPGRLESLRTADVRA 77
DB 88 VAALIAAGGESA-----PPGAASAPSPVPAEIVVAVPTPVYAEFTSLTAIVEQVELRP 140
QY 78 QVGGIIQKRLFOEGSYVYAGQPLVQIDSTYEANLESRAQOLATAQATLTAKADLARYK 137
DB 141 RVAGYIGQVTVPEGRVLEKGGQQLFLIDRVFKKAQDARARLRBAALALAFHEHBAE 200
QY 138 PLVAEAIVSRQEIADAVTAKSAAEGVKAQAQAIKASGININRSRITTPISGFGOSKYS 197
DB 201 LLVARKVVARERLDSATASRNASKAQVDAAARAALDAQLDGFTRVTAFIGRGVGHIOVT 260

```

```

QY 198 EGTLLMAGDTVLATIRQTNPMYV--NVTOSASEVMKLRQIAEGKLLAADGVIAVGIRK 255
DB 261 EGNVTVNG-VTALNISIVSDPLVYFVDERTYIQLALAPTRGRSEQAPR---KVALLLT 316
QY 256 DDGTVYPEKGRLLPADPVNVESTGQITLRAAVNDQNIIMPGLYRVLDQVAVDNAFV 315
DB 317 DES--YGRSSRLDPLANNAADRGTGTVRAVAVNDPDGQITPGLPAKVLETGKPPRAQVIV 374
QY 316 PQAV-TRGAKDTMIVNAQGMREPREVVAQOGGTNMTVTSGLKDGKVVVEGSIAGI 374
DB 375 ADHSIGTDQGRRYLVVDENKKTQYRVELSPMVDGLRVAVQGLQPGERIIVVKG---V 430
QY 375 TGAKKVTPK 383
DB 431 RPDQITPR 439

```

```

RESULT 12
US-09-252-991A-24485
; Sequence 24485, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196,136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24485
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24485

```

```

Query Match      13.7%; Score 277; DB 4; Length 394;
Best Local Similarity 27.1%; Pred. No. 8.4e-19;
Matches 104; Conservative 71; Mismatches 141; Indels 68; Gaps 14;

```

```

QY 14 LAAVALVLSGCG-KGGDAAGGQAPGRAPVAVGVVTHPQTVALTVELP----- 64
DB 28 LAAVIAVAVAILIAGIKVYSIRO--QIALPSAKPEPTIV-----TASIAEKRPWQSRLLPAI 79
QY 65 GRLESLRTADVRAQVGGIIQKRLFOEGSYVYAGQPLVQIDSTYEANLESRAQOLATAQA 124
DB 80 GSLKAPQGVTLTAIVSGTVADVFLPSGDQYKLDQPLQLBESDVEATLR-----TREA 132
QY 125 TLAAVADLARYKPLVAEAIVSRQEYD--AAVTAKSAAEGVKAQAQAIKASGININRSR 182
DB 133 DLGIARAAYQGRRLISKAISKSEFDRLLAQMKTSAIV-----AELKAA---LAKKR 183
QY 183 ITAPISGFIGOSKRYSEGTLLMAGDTVLATIRQTNPMYVAVTOSASEVMKLR-QIAEGK 241
DB 184 VLAFAAGTIGIRQVDVDVSPG--TPIATLQDLSILLPLFHLPEQDFPLLSRQOLVKVR 241
QY 242 LLAADGVIAVGIRKPDGTVPYPEKGRLLFADPVNVESTGQITLRAAVNDQNIIMPGLYV- 300
DB 242 VAAVPGQV-----FD-----AEIAINPKVDNETRNIQVRALENPRGKLLPGMFRAN 288
QY 301 -----RVLMDOQAV-----DNAFVVPQAVTRGAKDTMIVNAQGMREPRETV 344
DB 289 LEVWLPEEQRVVVVPETAIFETLYGDSIYVVGQKDEQGVSKDKGQPOOVRERRFRVRI 348
QY 345 AQOOGTMIYTSGLKDGKVVVEG 368
DB 349 GERREGIAVVEGEGEQQVVTSG 372

```

RESULT 13
US-09-328-7897 Application US/09328352
Sequence 7897, Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 7897
LENGTH: 452
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-7897

Query Match 13.5%; Score 272.5; DB 4; Length 452;
Best Local Similarity 23.1%; Pred. No. 2,9e-18;
Matches 99; Conservative 62; Mismatches 127; Indels 141; Gaps 11;

QY 65 GRLESLRTADVRAQVGIIQKRLFOEGSVYRAGOPLYOIDSSTYF-----ANLESARA 117
DB 60 GTLDATFKLISVGAQVSGVKMYVQLGDQVKGOLIAQIDSTTOENSLKTSIDANIKMLEA 119
QY 118 QLATQAATLAKADADLARYKPLVAABEAVSROEYDAVTAKRSGAAGVKAQAQAIKISAGI- 176
DB 120 QRLQQTASINEKOLEYRROQMYAQDATPRADLESABAAKYKQAQVRLADQIISAKIT 179
QY 177 -----NINRSRTAPISGFIGOSKVSSEGTLLNAGDTTVLATIRQTNPMYVNTQSASEV 230
DB 180 RSTAQTNGYTRIVAPDTGTVAIVTEBQTVNA-----NQSAPTVIKAKLQN-- 228
QY 221 MTLRQIABGKLADGVIAVGIRKDDGTVPYRPEKRLPADPVNVESTQITLRAAVEND 290
DB 229 MTKAQVSEADIMKV-----EKQOQVYFTLLGDETRKRYATLRQIEPAP 271
QY 291 QNI-----LMPGLYVRVLM-----QVAVD-----NAFVYPCQAV 320
DB 272 DSISESNSSTSTSSAVYNAALPDVPNTDGLRIDMTAQYIYVLSAKALLVPSSNL 331
QY 321 T-----RGAK-----DTVM 329
DB 332 SSKQPSGKQKQGSADKASSTPSABRKHQNGARLERLNLTPKQKQIIEQKATLSYVR 391
QY 330 IYNAQGGMEPREVYTAQOQGTWVIYTSGLKDGKVVYBESIGAITGAKKTPKEMASSE 389
DB 332 VLAQGGTTPQIILVGINNRVNAQVLAGIKGDQVVIAD-----SSB 433
QY 390 NOAAPQSG 398
DB 434 NSASANSNG 442

RESULT 14
US-09-252-991A-20610 Application US/09252991A
Sequence 20610, Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 20610
LENGTH: 399

TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20610

Query Match 13.2%; Score 267; DB 4; Length 399;
Best Local Similarity 25.6%; Pred. No. 8.2e-18;
Matches 93; Conservative 64; Mismatches 170; Indels 36; Gaps 8;

QY 11 AAALAAVALVSSCGKGDAAQGGPAGREAPAPVGVTVHPQVATLTVELPGRLESL 70
DB 38 SAALICAAVIGIAYVATGSAKKDGAFAG--YYPVKYALASVERVPRVDSGVELEAG 95
QY 71 RTADVRAQVGIIQKRLFOEGSVYRAGOPLYOIDSSTYFANLESARAQATLAKAD 130
DB 96 ROVVAABEAGRIIRIFAFESGQVQOQQLVLQVNDVAGQAEIRLKLQRLNABEILHARA- 154
QY 131 ADLARYKPLVAABEAVSROEYDAVTAKRSGAAGVKAQAQAIKISAGININRSRTAPISGF 190
DB 155 -----RKLVRNVAASOEQLDNAAVARDMALGAVRQTOALI-----DQKAIAPFSGQ 201
QY 191 IGOSKVSSEGTLLNAGDTTVLATIRQTNPMYVNTQSASEVVKLRBQIABGKLADGVIA 250
DB 202 LGIRRVHLGQYLVGAEPVASLVDAKTLKSNFSLDESTSPELKQOP-----LE 249
QY 251 VQIKFDDGTVPYRPEKRLPADPVNVESTQITLRAAVENDNIIIMPGLYVRVLMDOYAVD 310
DB 250 VLVDAYPGRSFP--ARISAIIDPLIGKSR-TVQVQALLDNPEGLLAAGMFASIRSRKADA 306
QY 311 NAFVYPCQAVTRGA-KDTVMIVNAQGGMEPREVYTAQOQGTW-----IVTSGLDQDGKVV 365
DB 307 PLSVPEETAVYTVAYGDTVFVAHQDGRPLSAKRSVIRIGERMDGREVEITLQGLAEGDRV 366
QY 366 VEG 368
DB 367 TSG 369

RESULT 15
US-09-252-991A-28543 Application US/09252991A
Sequence 28543, Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 28543
LENGTH: 406
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28543

Query Match 12.7%; Score 256; DB 4; Length 406;
Best Local Similarity 26.9%; Pred. No. 1e-16;
Matches 100; Conservative 68; Mismatches 154; Indels 50; Gaps 11;

QY 10 RAALAAVALVSSCGKGDAAQGGPAGREAPAPVGVTVHPQV-----ALTVELP 64
DB 39 RVALPAILICAGLLVCGG-----APPPEEHVRAVLAQVVKMAEFASATST 83
QY 65 GRLESLRTADVRAQVGIIQKRLFOEGSVYRAGOPLYOIDSSTYFANLESARAQATLAKA 124
DB 84 GDICARVQADSPFVGGKIVYRVLVDGVHVAAGVTLARLDQDQDSNVENNAQVAAQQA 143
QY 125 TLAKADADLARYKPLVAABEAVSROEYDAVTAKRSGAAGVKAQAQAIKISAGININRSRT 184


```

Db      144 QSKLADLNYQROKALLPKGYTQSEYDQALASVRSASQSLKAQAQOLANARDLLSYTEIR 203
Qy      185 ADISGEIGSKVSEGLTINAGDTVTIATIRO--TNPVNVYVYQS--ASEVMKLRRQIAEG 240-
Db      204 ASDAGVITRQAQEVGVQA--TVPIFLTARDGERDAVNVYESLFSDHVDQGRITVS-- 259
Qy      241 KLLADGVIAVIRKDDGVVPEKGRLLFADPVVNESTGQITLRA--AVPNDQITLMPG 297
Db      260 -LLGKREYVA-----SGKVRITPTVDERSGTLLKVKGLDVSFPAEMSL---G 302
Qy      298 LVYRVLMQVAVDNAFVVPQQAIVTR-GAKDTVMINAGSGMREYVTAQOOGTMIYTS 356
Db      303 SVNASSVAAPA-EHSEVVLFWMSLSKVGQBPVWMLDQCKARLQPVFARVASEKVIDG 361
Qy      357 GLKDGDKVVEG 368
Db      362 GLEAGQTVTVTG 373

```

```

RESULT 16
US-09-252-991A-21202
; Sequence 21202, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21202
; LENGTH: 389
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21202

```

```

Query Match      12.2%; Score 245.5; DB 4; Length 389;
Best Local Similarity 22.0%; Pred. No. 1e-15;
Matches 90; Conservative 72; Mismatches 189; Indels 59; Gaps 8;

Qy      4 YAFKMRRAALAAALVLSGCGKGDAAQGGPAGREAPAPVGVTVVHPQVIA-LTVE 62
Db      22 HMYRIPIPLVALSLFSSLLFACGNGTTPPAARPA-----IVQOPAGEVSA 70
Qy      63 LFGRLSELRADVAQVGGIIQKRLFOEGSVYRAGOPLYQIDSSSTYEANLESARAQLATA 122
Db      71 FFGRIARHHEBELAFRIGKTYIRLVEGKVKDQPLAELDPQDVRLQLEPAARQVSA 130
Qy      123 QATLAKADADLARYKPLVAEAVSROEYDAAVTAKRSAGVKAQAQAIKSGINLNSR 182
Db      131 EANIQTVAEYRRTLLDRMLVSHSGFENIONSVRAGEARLKQIRAFENVADNOAGYAV 190
Qy      183 ITAPISGFIGSKVSEGLTINAGDTVTIATIRQINPMVNVYVYQSASEVMKLRRQIAEKL 242
Db      191 LRSPDQGVIAARRVEGVQAQGVTFSLADGEREVLIGLPEHSFERFRIGOPVSVELM 250
Qy      243 LAAGVIAVGIKFDGTVYPEK---GRLFPADPVNVESTGO-----ITLAAVNDQNI 293
Db      251 SQDRRRFAGHIRELSPADPOSRTFAARVAFDRAITPDELQOSARVYVAAEAEP----- 305
Qy      294 LMPGLYRVLMQVAVDNAFVVPQQAIVTRGAKDTVMINAGSGMREYVTAQO----- 347
Db      306 -----LSVPLSALTAEAGQAFVWV-----VEPGSSTLRQAQVTRGP 341
Qy      348 -QGTNMTVYSGLKDGDKVVEGSIAGITGAKYTPKEMASSENOAAPQ 396
Db      342 YAEDRVVPVLEGLKAGDMVAVATGVV--LRGQVQRPIDRANRYVLAKE 389

```

```

RESULT 17
US-09-252-991A-28510
; Sequence 28510, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28510
; LENGTH: 770
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28510

```

```

Query Match      11.8%; Score 237.5; DB 4; Length 770;
Best Local Similarity 24.6%; Pred. No. 1.8e-14;
Matches 92; Conservative 61; Mismatches 176; Indels 45; Gaps 10;

Qy      8 AMRAALAA--AVLVLVSSCGKGDAAQGGPAGREAPAPVGVTVVHPQVIA-----AL 59
Db      414 AMKPFSLAGLFGFALLISGCG-----DEPPAPRPVLYTVVTKLKNDDIGRF 461
Qy      60 TVEIFGLSELRADVAQVGGIIQKRLFOEGSVYRAGOPLYQIDSSSTYEANLESARAQL 119
Db      462 AGSIQAVESV---LGFRTNGRIASRLFDVGDVGVGALLATIDPDDQOQLAASQDL 517
Qy      120 ATAQATLAKADADLARYKPLVAEAVSROEYDAAVTAKRSAGVKAQAQAIKSGINLNL 179
Db      518 ASBAQLIDAQANARQBELFARSVTAQARLDARTLKTQSAQSFQDAKAIVQARQDLS 577
Qy      180 RSRITAPISGFIGSKVSEGLTINAGDTVTIATIRQINPMVNVYVYQSASEVMKLRRQIAE 239
Db      578 YTRLVTDFDGYITTHMAEQVNSAQO---AVVTLARP-----EVREAFFDLPTVEAE 627
Qy      240 GKLLAADGVIAVGIKFDGTVYPEK---KGRLLFPADPVNVESTGQITLRAAVPNDQNIIMP 296
Db      628 S--LPADARPLVSAQLD-----PQARTGSIRELGPADASTRRRRVRLSLAQTPPEARL 680
Qy      297 GLYRVLMQVAVDNAFVVPQQAIVTRGAKDTVMINAGSGMREYVTAQOOGTNIYV 355
Db      681 GSTIQVQLSSAGSVRSVLPASVLLERDKTQVWVVDKQSSVALREVQVLSRDERQVIG 740
Qy      356 SGLKDGDKVVEGI 369
Db      741 QGLADGDRVRVAGV 754

RESULT 18
US-09-328-352-8183
; Sequence 8183, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 8183
; LENGTH: 500
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-8183

```

Query Match	10.9%;	Score 220;	DB 4;	Length 500;
Best Local Similarity	22.7%;	Pred. No. 4.9e-13;		
Matches	72;	Conservative	57;	Mismatches 112; Indels 76; Gaps 10;
Qy	71	RTADYRAQVGGIIQKRLFOE--GSYVRAGOPLYOIDSSTEYANLESARAQALATQAATLAK	128	
Db	132	QVAILQTRASGFVQ-RVYGHVAGDMVTQGSPIADISIPEMTGEOTEFIALVTRGDRSL--	188	
Qy	129	ADADLARYPELVAAEAVEROEYDAATYAKSAEAGVAAQAATISAGINLN-----	179	
Db	189	-----IOASHQRQLQLLGIPOVINQVERTH	213	
Qy	180	---RSRTAPISGFIGOSKVSSEGTLLNAGDTVIATIRQNPVYVNTQASAEWMKLR	235	
Db	214	RVSQNMNTLSAPVSGFIDSLSEVRNGMALAMQOT--LVTIKGISPLWLEASVEAQTACIKR	271	
Qy	236	QIAEGKTLAADGVIAVGIKEDDGTV-YPE--KGRLLFADPVNVESTQITLRAAVPNDON	292	
Db	272	-----GMKEATEFVAVPQIVSGKVIDILPLTIDSTSRITKRIEFLPNREG	315	
Qy	293	ILMPGLYRVLMDQVAVNDNAFVVPQQAATVTR-GAKDTMIYVNAQSGMEPREVTYVAAQOQSTN	351	
Db	316	MKPGFASFVAKESNNPQSN-LVIPQAIVIRGTGRNVVIVGHEQGFEPFVVVQLQSDGSK	374	
Qy	352	WIVTSGLKXDGKRVVVEG	368	
Db	375	IALQLGLKAGQKVVISG	391	
RESULT 19				
US-09-328-352-5003				
Sequence 5003; Application US/09328352				
Patent No. 6552958				
GENERAL INFORMATION:				
APPLICANT: Gary L. Breton et al.				
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER				
TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS				
FILE REFERENCE: GTC99-03PA				
CURRENT APPLICATION NUMBER: US/09/328,352				
CURRENT FILING DATE: 1999-06-04				
NUMBER OF SEQ ID NOS: 8252				
SEQ ID NO 5003				
LENGTH: 367				
TYPE: PR				
ORGANISM: Acinetobacter baumannii				
US-09-328-352-5003				
Query Match	10.7%;	Score 216.5;	DB 4;	Length 367;
Best Local Similarity	24.7%;	Pred. No. 6.6e-13;		
Matches	93;	Conservative	63;	Mismatches 165; Indels 55; Gaps 14;
Qy	40	RAAP-----APVGV-----VTVHPQYALTYELPGRLESRTADYRAQVGGIIQKRLFOEG	91	
Db	25	KEAPKTEERPYVWVTPSTLHEOK-----SYAGDVQARQOTALAFRVGGQVTRARYVDVG	79	
Qy	92	SYVRAGOPLYOIDSSTEYANLESARAQALATQAATLAKADADLARYKPLVAAEAVEROEYD	151	
Db	80	DEVRVGGVIAKLDEVEDAQLQNLNAKAKQLDNAQAQAKTASDELKGFQQLPINAVERSCYD	139	
Qy	152	AAVTAKRSAEAGVKAQAQAIKSAGINLN--SRITAPISGFIGOSKVSSEGTLLNAGDTT	208	
Db	140	---TYNQVDAQAALQOARSNVEVSANQGVNQLVSNKGVITFARNIEIGQVVAAGQAA	196	
Qy	209	VLAITRQTNPMYVNTQASAEWMKLRRLQIAEGKTLAADGVIAVGIKEDDGTVPE-----	263	
Db	197	VQALADGEREVVIGVEQA-----VSEIKVGAQAMI-----TLMSKPERF	237	
Qy	264	KGRLLFADPVNVESTQITLRAAVPNDONILMPGLYRVLMDQVAVNDNAFVVPQQAATVTRG	323	
Db	238	AGYVAVESPAADQSR-TFTYKVALKEGQSAIQLGQSAKRVFPSSQT--NMSVPLSSVS--	293	
Qy	324	AKDT---VMIYVNAQSGMEPREVTYVAAQOQSTNIVTSGLKDGKRVVVEGISTAGITGAKV	380	

[illegible]

Query Match	9.1%;	Score 184.5;	DB 4;	length 367;
Best Local Similarity	25.3%;	Pred. No. 9.2e-10;		
Matches	85;	Conservative	46;	Mismatches 114; Indels 91; Gaps 13;
QY	49	WYTHPQVATLVLPGRLESLRTADVA-----QVGIIOKLPFGESYVRACQ	99	
Db	40	IVLVSIIMLKVITLPSSV-VKTDADAVVEYTIAPKVGNIIEIYIKKHQTVKQQL	97	
QY	100	LYQIDISTYEANLESARAQIATAQATLAKA-----DADL-----	133	
Db	98	LARIADADYQALALEAESNVYAKAQAADINEAMLAVERQPTVIRETEAQLRKVEAGIKLTKD	157	
QY	134	--AAYKPLVAEAENVSR-----QEYDAATYAKRSA	160	
Db	158	NTAAYEQLALGASRLLITQOSKTTTPEYQVADLDSKKEVYDAQYOLNQYKIQVQAK---	214	
QY	161	EAGYKAAQAAIKSGININRSRTIAPISGFIQSGKVSSEGTLLNAGDTTVLATROTNDMY	220	

```

Query March 23.1%; Score 184; DB 4; Length 423;
Best Local Similarity 93.5%; Pred. No. 1.3e-09;
Matches 92; Conservative 66; Mismatches 178; Indels 56; Gaps 15;

QY      12 AATAAAV--ALVYSSCGKG-----GDAACGCGPAGREAPAPVGVVTHPQTVALT 60
      13 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      36 AATTAATIGLLVLTNSKNKKSNSSEPSGEGHGAEEBGEHDEGKPLT--LTAQOQWQONLK 94
QY      61 VE-----LPGRL--ESLRTADVAVOAGIIOKRLFOEGSSVVRAGOPLYQIDSS 106
      12 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      95 IEQAEIGEVQLOLTPYPAKLVTNTDROAHVSPSSGREAAYVELIGQVKKGQAL---AS 150
QY      107 TYEANTESARAQATATQATLAKADADLARYKPLVAABEAVSRQEYDAAVTKRSAEAGVKA 166
      15 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      151 LTVPDLTDQOQANQIOAOSNLIELARODYERERSLMSQGISAKODYQRAYNAYVQQAQIQVKA 210
QY      167 AQAARKAGI---NLNRSRTAPISGFIQGSXVSEGTLLAAGPTVATATROTNPVYN- 222
      13 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      211 SRSLTAFGAGSSSAGRYTLTAPLAGVSKKDVIAGENVALDQ--LFTINQDLMLHF 268
QY      223 VTQASAEVTKLRQRLQAEGLKLLAADGVIAVGIKFDDGTVPYEKGRLLFADPVNESTGQIT 282
      12 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      269 ILSNNANINQVPONQOIEFKSL-----QTGTTFSAQOVQSLTTE--ADAQTRQLQ 314
QY      283 LRAVPNDQNTLMPGLVRLMDQVADNAFVVPQOAVTR--GAKDPTM---IYNAQSGME 338
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      315 VRKAVTLANSELRLPMLMVAWVELNAGSTQKTRKKAQAQVQVEGKDVIPTFKIYKTGEFE 374
QY      339 PREVTAAQO-QGTNWI-VTISGLKDGDKRVVEG 368
      13 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      375 PVTVQLGQRSKDQGWVEVVKGINPSORXYIAEG 406

RESULT 24
US-09-252-991A-31115
; Sequence 31115, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18

```

; PRIOR APPLICATION NUMBER: US 60/074, 788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094, 190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 31115
 ; LENGTH: 517
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-31115

Query Match 9.1%; Score 184; DB 4; Length 517;
 Best Local Similarity 23.9%; Pred. No. 1.8e-09;
 Matches 79; Conservative 52; Mismatches 162; Indels 38; Gaps 8;

QY 51 TYNPCTVATLVLPG--RLESLRTADVRAQVGGIIQKRLFGSGSVYRAGQPLXQIDSTRY 108
 DB 195 TARTISLQVNSLPEIRNEDRTAHIVRPLPGIVDSVPANIGAKOGLAVISS--- 251
 QY 109 EANLESABAQLATAQATLAKADADLARYKPLVAABAVSROEYDAVATKRSAAQVKAQ 168
 DB 252 -POLSDQSEFAAQRRLSLAQSTYKREQLMKEGISAQEPFLARQGLQAEFALNNAR 310
 QY 169 AAIKSAQINL-----NRSRTAPISGFIGQSVSGCTLINAGDTTVALITRQTNPMYVNV 223
 DB 311 AKIALGQNPISLQGNRYELRAPFAGVLVEKHLTQCEPVDG--TANVFLLSLSVWATF 368
 QY 224 TQSAEWMKLRQIAEGKLLADGVIANGIKFD-----DGVVPEKRLIFADVNVMS 277
 DB 369 NVPA-----QLL---GGVRGSKVKVLAQALDSEV---EGTVSYIGLLGEQ 409
 QY 278 TGOITLRAAVPNDQNLMPGLVYRVLMQVAVDNAFVVAPOAVTRGAKDTVAVNAQGM 337
 DB 410 TRAAIARVTLSPSTWTRGLFVSQVVAETREKVLTVADGAVQVDGDEVFVFAADCF 469
 QY 338 EPREVTVAQOQGTNNIVTSGLKDGKRVVVEG 368
 DB 470 VVQPVKLGISDQGRVEVLEGLRAGSQVAASG 500

RESULT 25
 US-09-252-991A-18501
 ; Sequence 18501, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; PRIOR FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074, 788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 18501
 ; LENGTH: 354
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-18501

Query Match 9.1%; Score 183.5; DB 4; Length 354;
 Best Local Similarity 24.6%; Pred. No. 1.1e-09;
 Matches 86; Conservative 41; Mismatches 116; Indels 107; Gaps 12;

QY 8 ARAAALAAVALVSSCGKGDAAQGGPAGREAPAPVGVVTVHPOTVATLVELPRL 67
 DB 26 AVVVALGALAFWLS-----RDHQQDALRIYXVNDI--REVQALFQPGR- 70
 QY 68 ESLRTADVRAQVGGIIQKRLFGSGSVYRAGQPLXQIDSTRYEANLESABAQLATAQATLA 127
 DB 71 -----VMQMAFDEGDAVSAGARLALIDPQRYEALAAQAQVAVQAQAEIA 115

QY 128 K-----ADADLARYKPLVAABAVSROEYDAVATKRSAAE 161
 DB 116 KLRGRLRQREITQAREALRQALATETERNFQGSGLMSGATSQRVDAARTADQAA 175
 QY 162 AGVRAQAQAIKSAQINLNRSTRTAPISGFIGQSVSGCTLINAGDT-----TVLATI 213
 DB 176 AGVENAKAALISQASEGFEKEDIAAAEARLAAQAQAAQATTALADTELAPSSGTVIARV 235
 QY 214 RQTNPMYVNVVQSAEWMK-----RQIAEGKLLADGVIANG-----IKFDGQ-TYYP 262
 DB 236 REPQSMV--VSQSAVYSLSLDKPVYVRAYVESDL---GRIAGQTVVVRVSDSEKVV- 288
 QY 263 EKGRLIFADP-----VNVESSTQITLRAAVP 288
 DB 289 -RGQIGFISPRAEFTPKVTETDRLTDLVYRLRVVIDETSDRLRQGM 337

RESULT 26
 US-09-328-352-7671
 ; Sequence 7671, Application US/09328352
 ; Patent No. 6562958
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary L. Breton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 ; FILE REFERENCE: GTC99-03PA
 ; CURRENT APPLICATION NUMBER: US/09/328,352
 ; PRIOR FILING DATE: 1999-06-04
 ; NUMBER OF SEQ ID NOS: 8252
 ; SEQ ID NO 7671
 ; LENGTH: 356
 ; TYPE: PRT
 ; ORGANISM: Acinetobacter baumannii
 US-09-328-352-7671

Query Match 8.9%; Score 179; DB 4; Length 356;
 Best Local Similarity 25.0%; Pred. No. 3e-09;
 Matches 74; Conservative 45; Mismatches 83; Indels 94; Gaps 11;

QY 70 LRTADVRAQVGGIIQKRLFGSGSVYRAGQPL-----YQIDSTRYEANLES----- 114
 DB 60 IROYSLAFESQGRLOKLVQDEGDKVOQVATINTNALQIOAKQAQKLAQOBAIVQ 119
 QY 115 -----ARQIATQATLAKADADLARYKPLVA-----EAVSROEYDAVATKRSAA 160
 DB 120 EVGARPEITQAKQALASQAHELDTKNLQRLQILVSTTGRAISQBELDYAKSNQSHA 179
 QY 161 EAGVRAQA-----AIKSA-----GINLNRSTRTAPISGFIGOS 194
 DB 180 DAAVERQANLELIKARKEDREATRAQEVYTKANLDLINVNLQALRSPVNA----- 234
 QY 195 KYSEGTLINAGDTT-----VLATIRQTNPMYVNVVQSAEWMKLR-----ROI 237
 DB 235 -VVRARLOEVDMDTTAQAVVTLALTNPKMIRVYVNEQDLSSIKKSGTAQIIRDAYPNOP 293
 QY 238 AEGKLLADGVIANGIKFDGTVVPEKGR--LTFADPVNVESSTQITLRAAVPNDQ 291
 DB 294 MNKTI-----GIYISVAETPKTVQTEIRITLVY-----EVRVYVADPNDQ 335

RESULT 27
 US-09-328-352-5910
 ; Sequence 5910, Application US/09328352
 ; Patent No. 6562958
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary L. Breton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 ; FILE REFERENCE: GTC99-03PA
 ; CURRENT APPLICATION NUMBER: US/09/328,352
 ; PRIOR FILING DATE: 1999-06-04
 ; NUMBER OF SEQ ID NOS: 8252

SEQ ID NO 5910
 LENGTH: 358
 TYPE: PRT
 ORGANISM: Acinetobacter baumannii
 US-09-328-352-5910

Query Match 8.6%; Score 174.5; DB 4; Length 358;
 Best Local Similarity 24.6%; Pred. No. 8.5e-09;
 Matches 88; Conservative 50; Mismatches 139; Indels 81; Gaps 11;

QY 1 MAFVAFKMAAALAAVALVSSCGKGDAAGGQGP-----GREAPAPVGVVH 53
 DB 5 LGFFLLIILIAILYALMALFNVSTDNAYVGAETQITQTSWGVQAQVILKDTQTVH 64
 QY 54 POTVALTELPGLTSLRTADVAQVGIIOKRLFOE---GSYVAGOPLOYIDSTYEA 110
 DB 65 RGDV--LVIRIDRDALKALQAELAK--AKQYKQTAANSSSLNSQVVRAD----- 114
 QY 111 NLESAPQLATPQATLAQADADLARYKPLVAEAVSRQDYDAATAKSAENG----- 163
 DB 115 EINSKAOVQAQADYDRAAELENNRAQLAASGAVSEKELTYAQSAVETAKAGLELAKAG 174
 QY 164 -----VKAAQAIAKSAGININRSRTAPISGFI 191
 DB 175 LAQATSSRKAESTLANEALIQVSETSPDVQAQHVQAQDLERTVIRAVPDGYI 234
 QY 192 GQSKVSEGLTNAGDTVLTATIRQTNPMYVNVVQSASEVMKLR--ROIAGKLLAADGYI 249
 DB 235 TRRNIVQQRVAPG--TSMNMIIVPLNDLYDANFESQKKVPRQPTLTLSLDYGDVE 292
 QY 250 AVG--IKFDGTIVYPEKRLFPADVNESTGQITLRAAVPNDONILMPGLYVRVLM 305
 DB 293 YHGKVVSGSGT-----GSAFALIPQONATGMWIKVQRLP-----VRIALD 334

RESULT 28

US-09-252-991A-25258
 Sequence 25258, Application US/09252991A
 Patent No. 6551795

GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 PRIOR FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 25258
 LENGTH: 510
 TYPE: PRT
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-25258

Query Match 8.6%; Score 174.5; DB 4; Length 510;
 Best Local Similarity 22.7%; Pred. No. 1.5e-08;
 Matches 99; Conservative 61; Mismatches 145; Indels 131; Gaps 19;

QY 8 AMRAAALAAVALVSSCGKGDAAGGQGPAGREAPVGVVHPTQVALTELPGRLE 67
 DB 102 AQAVQALGLRTAEVH--GKIG-----ADVAVVGTVLINBRDYSI----- 139
 QY 68 ESLRTADVAQVGIIOKRLFOE---GSYVAGOPLOYIDSTYEAESARAQATAQAT 125
 DB 140 -----VQARTAGFVE--RVYARAAGDVIAAGAPL----- 166
 QY 126 LAKADADLARYKPLVA-AEAVS--RQDYDAATA---KRSAGV--KAAQAIAKSAGINLN 179
 DB 167 -----ADLLRPKWTIAQREFLSVAMGDAPLTAARQLLLILGMPQALLAQVETGEPKG 221

QY 180 RSRITAPISGFIQSKVSEGLTNAGDTV---LATI-----RQTNPMYVNTQSA 227
 DB 222 LYVTTPQGGIVAEILMVRQGWTVSAGESLVNGLASVIAEAVPEAGSGPLOG----- 276
 QY 228 SEWKKLRQIAEGKLLADGVIAVGIRKFDGTVPEKRLLPAPVNVESQITLRAAV 287
 DB 277 -----QTAQVLAAPFGEVL-----QARIVSIIPENARDRTTVKVRLEM 315
 QY 288 PNDONILMPGLYVVLMDQVAVDN---AFVPOQAVTR--GAKDTVMIVNAQGMPEPREV 342
 DB 316 ANPQRLKAG-----MSGQIALKNEQRPALLVPESEAVIRCKRLALAYVDSRGFHPREV 370
 QY 343 TVAQOQGTNMTVITSGLKDGDRVVEG-----ISAGITGAKVTPEKMASSNQAAAP 395
 DB 371 QVGAEEVGQVLVNVNGLLAAGQGVVSAQFLIDSEASLGVLP-----GAASAPASP 422
 QY 396 QSGVQTASEACTASEA 411
 DB 423 HNGHGTSSPTAAATNA 438

RESULT 29

US-09-252-991A-23319
 Sequence 23319, Application US/09252991A
 Patent No. 6551795

GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 PRIOR FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 23319
 LENGTH: 495
 TYPE: PRT
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-23319

Query Match 8.6%; Score 173.5; DB 4; Length 495;
 Best Local Similarity 20.5%; Pred. No. 1.8e-08;
 Matches 97; Conservative 60; Mismatches 143; Indels 173; Gaps 15;

QY 23 SSCGKGDAAGGQGPAGRE---APAPVGVVHPTQVALTELPGRLE---SLRTADV 75
 DB 48 ATAGRRTPAQGPGEPRADRSFRPAGPGRFPTRFPLRCARYRESAGRMNEAPMTTANS 107
 QY 76 RAQVGGIIOKRL----- 88
 DB 108 ETPAGNPKRRKWLILLAVILATLASVAMEFFYGRWHEDDTDAYINGNVQITPQIVGT 167
 QY 89 -----QEGSVVAGOPLOYIDSTVE-----ANLESARAQLAT 121
 DB 168 VSIIGADGDLVRKQGLVRFPSDDADIALQREANLAHTVROVRLFSNVQGRVRAEVAT 227
 QY 122 AQATLAKADADLARYKPLVAEAVSRQDYDAATAKSAENG----- 163
 DB 228 RKVALAQAEDYKRRKVLADGAIISQELAHARDALDSAKASLTSSEQOINTNRALVDOT 287
 QY 164 -----VKAAQAIAKSAGININRSRTAPISGFIQSKVSEGLTNAGDTTV----- 210
 DB 288 QITSHDPKMAAAQRLQAVYDDARSTIVAPVTGVARSQVQGRVQPGNALMAVVELDQ 347
 QY 211 -----ATIRQTNPMYVNVVQSASEVMKLRQIAGKLLAADGVIAVGIFPDGTVYPEKGR 266
 DB 348 ITWDANFKETQLKMRIGQP---VEIRSDLYSGDVARYSGTVBSLGV---GT-----GS 394
 QY 267 LRFADPVNVESTQITLRAAVPNDONILMPGLYVRVLMQVAVDNAPVPOQAVTRGAK- 325

```

Db      395 AFSLLPAQATNGNMWIKYQVRP-----VRHID-----PQELQHPRLI 433
QY      326 ----DTVMIVNAQG-----MEPRE-----VTVAAQOOGTMMIVTSGKODKRV 365
Db      434 GLSMVKKVDLHDQSGPAAQCPREALFSTDVYQOO-----LASADKLI 477

RESULT 30
US-09-328-352-7412
; Sequence 7412, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7412
; LENGTH: 365
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7412

Query Match      8.6%; Score 173; DB 4; Length 365;
Best Local Similarity 22.6%; Pred. No. 1.2e-08;
Matches 67; Conservative 55; Mismatches 104; Indels 70; Gaps 10;

QY      75 VRAOVGGIIOKRLFOEGSYVRAGOPLYOIDSSTYE-----NLE 113
Db      72 VMPKISGEVLEILLINDNQVKKGETLAVLDHRDYQARHYDAHVSLSKALGVQOQNEK 131
QY      114 SARAOATATQATLAKADADLA-----RYKPLVAABAVSRQEDYDAAVTAKSAGYKA 166
Db      132 SARSSITTEANGSVAAQADLARKKEFERYOPLDKGVTRKONFEGISQYITLADQOLSK 191
QY      167 AQA-----IKSAGINLN-----RSRTAPISGFIGOSKYSE 198
Db      192 AQAANAABEAQGLSQASRAQLADIQSANANLNYVDLASSKVSFVSGKISGLATQK 251
QY      199 GTLLNAGDTTLATIRQINPMYVNTQASAEVWKLRRQIAEKKLLAADGVIAVGIFPDG 258
Db      252 GSRVSP-QTRLMALIIPE-NSLYQANFKETQLEKM--HIGQYKCLKDAVPSLNFQKIE 307
QY      259 TYPEKGRLLFADPVVNESTG-----QITLRAAVPNDQNI--LMPGLYVRVLM 305
Db      308 SFPSPASGATFSLMPDN-ATGNFNKVVQRIPIVRIADSSPHIDLKPKMSVSATVD 362

RESULT 31
US-09-328-352-4722
; Sequence 4722, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4722
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4722

Query Match      8.5%; Score 171; DB 4; Length 386;
Best Local Similarity 22.9%; Pred. No. 2.1e-08;
Matches 83; Conservative 56; Mismatches 120; Indels 104; Gaps 13;

QY      37 PAGREAPAPVGVVTVHPQTV--ALTVELGRLESL-----RT 72

```

```

Db      31 PAPAPAPASKLIPTKRSSTLLMGLVILITVILWAMRIGPATSVQOQDINSYVKGKT 90
QY      73 ADVRAOVGGIIOKRLFOEGSYVRAGOPLYOIDSSTYE-----ANLESARAOATATAQAT 125
Db      91 TILSQINGVYKVDLVKDFHVKKGOVLMHIDATTTIDQKVAQAASGVQAKITLANQOS 150
QY      126 LAKADAD-----LARYKPLVAABAVSRQEDYDA-----VTAK 157
Db      151 IAQKQADIVAAQAKVEQVRAQYELSLAQRLRYQOLGNSGAAKSQEDPRAADAENNLAL 210
QY      158 RSEANGVAAQAALIKSAGI-----NLNSRTAPISGFIGOSKV 196
Db      211 KQAEANVIVAAKEALKTAQVABEAGLEAQSAAQLDQATTQKDYSAVAPMDGQGEVNP 270
QY      197 SEGTLNAG-----DTVLATIRQINPMYVNTQSA--SEVMKLRRQIAEKKLLA 244
Db      271 RVGQVVAAGSOLLVILPQOTVIANFKETQIANRRIGQKAMFYDAMKHKFTGHVEQIS 330
QY      245 ADGVIAVGIFPDGTYPEKGRLLFADPVVNESTGQITLRAAVPNDQNI--LMPGLYVRV 302
Db      331 P-----AAGSEF--SVLKPDNATGNFT--KVQRIAVRITID--PNOEGIEHLRPGMSVIT 380
QY      303 LMD 305
Db      381 SVD 383

RESULT 32
US-09-328-352-8202
; Sequence 8202, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 8202
; LENGTH: 336
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-8202

Query Match      8.3%; Score 168; DB 4; Length 336;
Best Local Similarity 22.3%; Pred. No. 3.3e-08;
Matches 69; Conservative 54; Mismatches 106; Indels 80; Gaps 10;

QY      75 VRAOVGGIIOKRLFOEGSYVRAGOPLYOIDSSTYEANLESARAOATATAQATLAKADADLA 134
Db      54 VSSDVAGLVTEVLVODNOTVKKGOVLFKIDVSRALDVEQAKSDLAKKAAYAAQQAQGLA 113
QY      135 -----RYKPLVAABAVSRQEDYDAVTA-----RS 159
Db      114 QAKANLJKSTNIKLAENKARYSNLMDG-AISQEOQOVFATRDQSHAEHQLOAAIQ 172
QY      160 AEAGVKAQAALIKS-----AGINLNSRTAPISGFIGOSKYSEGLTLNAGDTT 208
Db      173 AEATIKQOQALVEATSNLHLAEINMHRRAVAAVAPADGTLSPFDRPGYVVGQAVAAAL 232
QY      209 -----VLATIRQINPMYVNTQASAEVWKLRRQIAEKKLLAADGVIAVGIFPDGTYVP 262
Db      233 DRKQLYVGVFEETKLDRIHIGDEATYQMDRQIKIGHVG-----IASGIEDERS--- 285
QY      263 EKGRLLFADPVVNESTGQITLRAAVPNDQNIIMPGLYVRVLMQVAVADNAFVPOQAVTR 322
Db      286 SSSKLL--ANVNPTFSWVRLAQRVP-----VKIYLDAPK--ELAFVS 325
QY      323 GAKDTVMIV 331
Db      326 GRTATVHHI 334

```

RESULT 33
US-09-252-991A-20336
; Sequence 20336, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 20336
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20336

Query Match 8.1%; Score 164.5; DB 4; Length 314;
Best Local Similarity 25.4%; Pred. No. 6.6e-08;
Matches 75; Conservative 41; Mismatches 118; Indels 61; Gaps 10;

QY 74 DVRAVGGIIOKRLFOEGSYVRACQPLVQIDSYTEANLESARQATATLAKADADL 133
DB 59 NVADVSGIVADVPRDQVLKQGLMLQIDPDHRIAVAKAESLVASRKATLMLROLN 117
QY 134 ARYPLVAEAASRQYDAVATKASAEAGVKAQAQAISAGINILNSRITAPISGFIQ 193
DB 118 ARRAEDMENVSHESRDANTAAAMADYEQAQADARLILERRVAAQVDGYVTN 177
QY 194 SKVSEGTLLNAGDT-----TVLATTQTNPMYVNTQSASEVMKLR 235
DB 178 LNVHGDVARVGEAKMAVIDKNSYVWVGYPETKLPYRREGDPYDM----- 223
QY 236 QIAGKLLAAD-GYIAVGIKDDGTVPEKGRLLFADPVNVESGQITLRAVPNDNITL 294
DB 224 QLMGSEHLKGVESIAKGIYDRDN---PE-SRELTAAD--VNPETNWRVLAQRVP----- 271
QY 295 MPGLVYRVLMQVAVDNAFVVPQAVTRGAKDTVIYVNAQSGMEPREVTVAQOQG 349
DB 272 -----VRVHIDE-----VPDGVLSAGITCTIVAKPGG--RDDQSAHAQAPG 311

RESULT 34
US-09-252-991A-17970
; Sequence 17970, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17970
; LENGTH: 357
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17970

Query Match 8.0%; Score 161.5; DB 4; Length 357;
Best Local Similarity 26.0%; Pred. No. 1.6e-07;

Matches 69; Conservative 40; Mismatches 101; Indels 55; Gaps 10;
QY 71 RTADVRA-----QVGGIIOKRLFOEGSYVRACQPLVQIDSYTEANLESARQATATQ 123
DB 107 RDAVRADVNVVADVSGVWTDLEVKDQNVKQVDVLMRIDQERQANLEQARRVAELRH 166
QY 124 ATLAKADADLARYPL---VAEAVSRQYDAVATKASAEAGVKAQAQAISAGINILN 179
DB 167 QOYLRRQNEAARRSRILGAIISADEKQENAI-----ARSEQEALQVKTALNLK 221
QY 180 RSRITAPISGFIQSKVSEGTLLNAGDTVIATIRQTN---PMYVNTQSASEVMKLRQ 236
DB 222 RSELRAARNGQVNLRLAQGNVATAGQ-AVALVDQGSFYVVAAYFEETKLPGRVGMRAQ 280
QY 237 IABGKLLAADVAVGIAKFDGTVYPEKGRLLFADPVNVESGQITLRAVPNDQ---NI 293
DB 281 V---RLMSGDQPI-----DGT-----VESISGITTDRNSTPDCQLLANV 316
QY 294 -----LMPGLVYRVLMQVAVD 310
DB 317 EPTFNWVRLAQRIPIVRIRLDQVPAD 341

RESULT 35
US-09-252-991A-23589
; Sequence 23589, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23589
; LENGTH: 399
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23589

Query Match 8.0%; Score 161.5; DB 4; Length 399;
Best Local Similarity 24.1%; Pred. No. 1.9e-07;
Matches 92; Conservative 54; Mismatches 144; Indels 91; Gaps 15;

QY 11 AALAAVALVLSGCGKGDAAQ-----QPA-GREAPAPVVG-----VV 50
DB 23 AAQOPAAALRRVAAGLARSAPVRVVLRLDRSDPVDGVDVPMNLQCKPRLATLLAV 82
QY 51 TVHPQVATLVELPGRLESRLTADV-----AOVGGIIOKRLFOEGSYVRACQPLVQI 103
DB 83 LVAAALVFLAMERYTRTPWRDARVADVTLTSADVGLITRLPLSNQVPEKELLSTI 142
QY 104 DSSITY-----ANESARAQATATLAKADADLARKPLVAEAVSRQ 149
DB 143 DPARYELAVLARAVVAVARPAALGESPAALIASQAQIRQRSEQRSAALQKSMLSVE 202
QY 150 YDAAVT-----AKSBAEGVKAQAQAISAGINILNSRITAPISGFIQSK 195
DB 203 NEKAQDVTLSAQABELRNQASLGLAQANVELAEHALAQADLERTVRAFPVSGVYTNLQ 262
QY 196 VSEGTLLNAGDTVIATIRQTNPMYVNTQSASEVMKLR-RQIAEGKLLAADVAVGIRK 254
DB 263 TRGDVYAHAG--VPLALVYDRDSFYVSGYFEETKLPGRIRGSRARVELMS-----GER 313
QY 255 FDDGTYTP-----EKGRLLFADPVNVESGQITLRAVP-----NDQNIYM 295
DB 314 F-DGRVQSIAMAITDRENAESRLAN--INPSYTWKLAQRIPVRIAIDPAVQRORNTLR 370

THIS PAGE BLANK (USPTO)

11-Jul-2000; 2000US-0614150.
(PEKE) PE CORP NY.
Venter JC, Adams M, Li PWD, Myers EW;
WPI; 2001-656860/75.
N-PSDB; ABL05732.
New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -
Disclosure; SEQ ID NO 11679; 21pp + Sequence Listing; English.
The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB22072).
The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

DR N-PSDB; AAH81443.
XX
XX Novel nucleic acids encoding proteins required for *Escherichia coli*
PT proliferation, useful for screening for antimicrobial agents -
XX
XX Claim 19; Page 554-555; 596pp; English.
XX
XX The present invention describes a purified or isolated nucleic acid
CC sequence (1) consisting essentially of one of the 93 nucleotide sequences
CC given in AAH81202 to AAH81294, where expression of the nucleic acid in a
CC microorganism is capable of inhibiting proliferation of a microorganism.
CC (1) have antibacterial and antibiotic activities, and can be used in
CC gene therapy. Expression of (1) in a microorganism inhibits proliferation
CC of the microorganism, and the manufactured antibiotic is useful for
CC reducing the activity or level of a gene product required for
CC proliferation of a microorganism in a subject, specifically humans. The
CC nucleic acids that inhibit bacterial growth or proliferation can be used
CC as antisense therapeutics for killing bacteria. In addition to
CC therapeutic applications, the nucleic acid sequences complementary to
CC sequences required for proliferation can be used as diagnostic tools.
CC For example, nucleic acid probes complementary to proliferation-required
CC sequences that are specific for particular species of microorganisms can
CC be used as probes to identify particular microorganism species in
CC clinical specimens. AAH81295 to AAH81487 encode the *Escherichia coli*
CC proteins given in AAH81295 to AAH81431, and AAH81488 to AAH81491
CC represent oligonucleotides, which are used in the exemplification of the
CC present invention.
XX
XX Sequence 355 AA;
SQ
Query Match 6.9%; Score 138.5; DB 22; Length 355;
Best Local Similarity 26.3%; Pred. No. 0.0011;
Matches 52; Conservative 32; Mismatches 55; Indels 59; Gaps 6;
OY 65 GRLESIRADYVAVGIIQKRLFOEGSYVRAGOPLYQIDSTYEANLESARAOLQA 124
DB 41 GRIETE-VDIASKAGRIDITLVKEGKVRGEVLAKMDTRVLOEQLALAQIKKES 99
OY 125 TLAKADADL-----ARYKPLVAEAVSRQEDY----- 151
DB 100 AVAAQAQLLEOROSFTRAQSLVNRQAELEDSVAKHRRSLAQRAISAQQLDDRAA 159
OY 152 --AATTAASAAGYKAAQAIAKSGIN-----LNRSRTAPIS 188
DB 160 AESAALAESAKAQAQVSAKAAIEAARTNIIQAQTRVEAQAQTERIAADIDDEIKABRD 219
OY 189 GFIGSKVSE-CTLNAG 205
DB 220 GRV-QRYVAEPGEVLAA 236
OY
DB
RESULT 38
AAVS5920
ID AAVS5920 standard; Protein; 444 AA.
XX
XX AAVS5920;
AC
XX
XX 15-FEB-2000 (first entry)
DT
XX
XX Pseudomonas fluorescens ABC transporter cassette component B.
DE
XX
XX ATP-binding cassette; transporter; operon; lipBCD; *Serratia marcescens*;
KW microbe; protein secretion.
XX
XX Pseudomonas fluorescens.
OS
XX
XX JP11276172-A.
FN
XX
XX 12-OCT-1999.
PD
XX
XX 27-MAR-1998; 98JP-0080597.
PF
XX
XX 27-MAR-1998; 98JP-0080597.
PR

XX
XX (TANA) TANABE SEIYAKU CO.;
PA
XX
XX WPI; 1999-626936/54.
DR
XX
XX N-PSDB; AAZ22700.
PT
XX
XX An ABC transporter gene - used to increase the ability of a microbe to
PT secrete proteins
PS
XX
XX Claim 2; Page 16-17; 28pp; Japanese.
XX
XX This sequence represents an ATP-binding cassette (ABC) transporter
CC component B from an *Pseudomonas fluorescens* strain 33 and encoded by
CC an ABC transporter operon. The operon comprises 3 genes where the
CC termination codon of the first gene overlaps with the initiation codon
CC of the second gene. The sequences of the encoded proteins have 60, 44
CC and 46% homology respectively to the lipBCD proteins from *Serratia*
CC *marcescens*. The novel gene and protein can give or increase the
CC ability of a microbe for secreting a protein.
XX
XX Sequence 444 AA;
SQ
Query Match 6.8%; Score 137.5; DB 20; Length 444;
Best Local Similarity 22.8%; Pred. No. 0.0018;
Matches 79; Conservative 65; Mismatches 103; Indels 99; Gaps 17;
OY 60 TVELPGRLESIRADYVAVGIIQKRLFOEGSYVRAGOPLYQIDSTYEANLESARAOL 119
DB 58 TVVSGKRKAVQTLS-----PGVSRILVREGEAVKQGPFRIDQONQADVHSLQAY 112
OY 120 ATQAATLAKADADLARYKPLVAEAVS-----RQYDAAVTAASAAGVK 165
DB 113 RMAVASVARQMSERDNOSTITFPALSGNPDALALVLEGRQLFSRREAFAREQAGIR 172
OY 166 AQAATISAGININ-----SRITA-----PI-SGFTQ-----SK 195
DB 173 --ANIGATRAQLQGMRRASDLTRAQOSLRDQNLNLPADNGYIPRNLMEXYROLQ 229
OY 196 VSEGTLLNAGDT-TVLATIRQTNPMYVNVTSASEVWK-LRQIAEGKLA----- 244
DB 230 VQDQLAQNTGSGVBEQALLESR---LKLQHSREYQKVRSLADQALSLTLEQQLTS 286
OY 245 -----ADGVIAVGIKFDGTYYPEKGRLLFADPVNVSTGQITLRAAVPNDQ 291
DB 287 AGFDLQSEINAPADG-IAVNL-----SVHTE-GAVVRA-----GE-TLEIYVQDT 330
OY 292 NIIMPGLYVRVLMQVA-----VDNAFVVPQQAATVRGAKDTVMIVNA 333
DB 331 RLEVEGRPLPVHLVDKVGTHLPVDILFTAFNQSRTPRVPGEVSLISA 376
OY
DB
RESULT 39
ABB61629
ID ABB61629 standard; Protein; 1201 AA.
XX
XX ABB61629;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX *Drosophila melanogaster* polypeptide SEQ ID NO 11679.
DE
XX
XX *Drosophila*; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
XX *Drosophila melanogaster*.
OS
XX
XX WO200171042-A2.
FN
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
XX
XX 23-MAR-2000; 2000US-191637P.
PR

XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 645 AA;

Query Match 7.5%; Score 151; DB 22; Length 645;

Best Local Similarity 26.5%; Pred. No. 0.0002;
 Matches 63; Conservative 43; Mismatches 96; Indels 36; Gaps 8;

OY 27 KGGDAAAGGQAPRPAAPVVG-----VTVHPQTALVLPGLRLSRTADVRA 77
 DB 361 RGPKHGQSPQ--NRAPFLDAGELSWRTDHIISIAFGKMFSTGENWSPFFRNFA--W 414
 OY 78 QVGGIIOKRLPOGGSV-----RAGGPVIOIDST-----YANLESARAOATQA 124
 DB 415 QWGEREEQAINNEGWIISPGEDTGNADKLNNNTAGAEINQKEVATLMELRAQQA 474
 OY 125 TLAKADADLARKYPLVAEAVSRQETDAVTT--AKRSAAG-----VKAQAARISAGIN 177
 DB 475 ELKARVYVSRQRLAQTAKVSCQDLDTNATMAKQAQIGTIDQIKRNOASLTATKTN 534
 OY 178 LNRSRITAPISGFIQSKVSG-TLLNAGDTTVALTIRQTNPMVYVNTQSASEVKLR 234
 DB 535 LDYTRIVAPMAGEVVIQITLQGGVTIAAQAQANILTLADMSMVLKQAVSEADVHTK 592

RESULT 36

ABG17838
 ID ABG17838 standard; Protein; 684 AA.

XX AC ABG17838;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #17829.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 XX food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.
 XX N-PDB; AAS82025.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
 XX diagnostics, forensics, gene mapping, identification of mutations
 XX responsible for genetic disorders or other traits and to assess
 XX biodiversity

XX PS Claim 20; SEQ ID No 48197; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 684 AA;

Query Match 6.9%; Score 140; DB 22; Length 684;

Best Local Similarity 53.6%; Pred. No. 0.002;
 Matches 30; Conservative 9; Mismatches 17; Indels 0; Gaps 0;

OY 90 EGSYVRAGPPIYQIDSTSTYENLESARAOATQAOTLAKADADLARKYPLVAEAV 145
 DB 102 EGSVDVQAGSLYQIDPATYQANVDSAKGELAKSEAAAAIAHLTVKRYVPLVAKNEV 157

RESULT 37

AGG98387
 ID AAG98387 standard; Protein; 355 AA.

XX AC AAG98387;

XX DT 21-SEP-2001 (first entry)

XX DE Escherichia coli protein sequence SEQ ID NO:435.

XX KW Escherichia coli; identification; proliferation; microorganism;
 XX antimicrobial; antibacterial; antibiotic; gene therapy; diagnosis;
 XX bacterial growth inhibition.

XX OS Escherichia coli.

XX PN WO200148209-A2.

XX PD 05-JUL-2001.

XX PF 19-DEC-2000; 2000WO-US34419.

XX PR 23-DEC-1999; 99US-0173005.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Forsyth RA, Ohlsen KL, Zyskind JW;

XX DR WPI; 2001-457376/49.

CC aluminum-resistant microorganisms. The products of the invention can be
 CC used to impart aluminum resistance to other organisms, particularly
 CC plants or microbes, especially to allow them to be grown in presence of
 CC aluminum at normally toxic levels, e.g. in acidic soils. This sequence
 CC represents the Acidiphilium cryptum Alub proteins described in
 CC specification.

XX Sequence 262 AA;

Query Match 7.8%; Score 158; DB 23; Length 262;
 Best Local Similarity 26.7%; Pred. No. 1.3e-05;
 Matches 67; Conservative 39; Mismatches 113; Indels 32; Gaps 8;

QY 7 KAMRAALAAVALVLSGCKGDAAGGQAPAGREAPVVGVT-VHPQTALVTELP 65
 DB 29 RRVRGAVLAALAV-----GVAALAIAGMTVG--APPVHVYVTKVGGDTTAVVTG 81
 QY 66 RLESRTADVRAQVGGITQKLFQSGSYVRAGOPLYQIDDSYEANLESARQALTAQNT 125
 DB 82 TVNPILTVYVGSVAGGIIQIDISCDVNTIVKLGQICARIDPRPYQAVVDQRAVAVDOAO 141
 QY 126 LAKADADI-----ARYKPLVAEAVSRQEDVAVTAKRSAAG-----VKAQAQAI 171
 DB 142 LKDEASLDYARAGNTRRLASLVPORYASENADSAKATTRQLEAOTLADRAVIRODEAKT 201
 QY 172 KSAGINLNRSRITAPISGFIGOSKVSSEGLTNAGDTTVLATIRQTNPMYVNTOSASEVM 231
 DB 202 AAAQANLGYTDIVSVNGIVSRAVNTIG-----QTVAAAP-QPTPLFLATDITR--M 251
 QY 232 KIRROIAGK 242
 DB 252 QVDIVNSEGDI 262

RESULT 31

ID AAU35553 standard; Protein; 390 AA.

XX AAU35553;

DT 14-FEB-2002 (first entry)

DE Haemophilus influenzae cellular proliferation protein #194.

KW Antisense; prokaryotic cellular proliferation protein;

KW antibiotic; antibacterial; drug design.

OS Haemophilus influenzae.

PN WO200170955-A2.

PD 27-SEP-2001.

PF 21-MAR-2001; 2001WO-US09180.

PR 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207727P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

DR MPI; 2001-611495/70.

XX N-PSDB; AAS53412.

PT New polynucleotides for the identification and development of
 antibiotics, comprise sequences of antisense nucleic acids -

XX Example 3; Seq ID No 11146; 511pp; English.

PS The invention relates to antisense inhibitors of genes essential to
 XX prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 390 AA;

Query Match 7.8%; Score 156.5; DB 22; Length 390;
 Best Local Similarity 21.6%; Pred. No. 3.2e-05;
 Matches 71; Conservative 51; Mismatches 115; Indels 91; Gaps 10;

QY 75 VRAQVGGITQKLFQSGSYVRAGOPLYQIDDSYEANLESARQALTAQNT 122
 DB 64 VSSQVAGNVAKINADNDKVAAGDILVELDNTNKLSEFOKSNLANAVRVEQLGFTVQ 123
 QY 123 -----QATLAKADADILARYKPLVAEAVSRQEDVAVTAKRSAAGVKAQ----- 168
 DB 124 QLOGAVHANETSLAQGNLRRVQLEKMGAIKDESFQHAKEAVELAKANINASKNQALAA 183
 QY 169 -----AAIKSAGINLNRSRITAPISGFIGOSKVSSEGLTNAGDTT 208
 DB 184 NQALLRNVPLEQPOIQINAINSLQAWMLNLRRTIRSPIDGVYARBNVQVQAVSGAL 243
 QY 209 VLATIRQTNPMYVNTOSASEVMKLRQI-----AEGKTLADGVYIAGIFDGTVP 263
 DB 244 MAVVSNEMQMLEANFKETQLTNMRIGOPVKIHFDLYGKKEPDEVIN-GIEMGTGMAF-- 300
 QY 264 KGRLLFADPVVNESTGQTLAAVNDQNIIMPGLYVRVLMQVAVNDAPFVPPQ----- 318
 DB 301 --SL--PSQNAAGNMWIKVQRPV-----VRIKLD-----POQFTETP 334
 QY 319 -AVTRGAKDTVMIVNAQGM-----EPR 340
 DB 335 LRIGLSATKVRISDSSGAMLRKTEPR 362

RESULT 32

ID AAE30466 standard; Protein; 390 AA.

XX AAE30466;

DT 24-FEB-2003 (first entry)

DE H. influenzae emrA protein (multidrug resistance protein A).

KW Virulence; veterinary; infection; pneumonia; antimicrobial drug; vaccine;

KW antibiotic; gene therapy; antibacterial; multidrug resistance protein A;

OS Haemophilus influenzae.

PN WO200277020-A2.

PD 03-OCT-2002.

AC ABP54530;
 XX
 DT 24-JAN-2003 (first entry)
 XX
 DE Actinobacillus pleuropneumoniae apvd protein SEQ ID NO:131.
 XX
 DE Actinobacillus pleuropneumoniae apvd protein SEQ ID NO:131.
 XX
 KW Antibacterial; vaccine; gram negative bacterial virulence gene;
 KM identification; virulence; Pasteurellaceae.
 XX
 OS Actinobacillus pleuropneumoniae.
 XX
 PN WO200275507-A2.
 XX
 PD 26-SEP-2002.
 XX
 PF 17-JAN-2002; 2002WO-US01971.
 XX
 PR 15-MAR-2001; 2001US-0809665.
 XX
 PR (PHAA) PHARMACIA & UPJOHN CO.
 PA
 PI Lowery DE, Fuller TE, Kennedy MJ;
 PI WPI, 2002-740868/80.
 DR N-PSDB; ABQ83530.
 DR
 XX
 PT New mutant gram-negative bacteria, useful as vaccines and for
 PT identifying new anti-bacterial agents that target virulence genes and
 PT their products -
 XX
 PS Claim 36; Page 296-297; 350pp; English.
 XX
 CC The present invention describes a gram-negative bacteria comprising a
 CC mutation in a gene, where the mutation results in decreased activity of
 CC a gene product encoded by the mutated gene. Also described is a method
 CC for producing a gram-negative bacteria mutant or an attenuated
 CC Pasteurellaceae bacteria. The mutated genes have antibacterial activity
 CC and can be used in vaccines. The gram-negative bacteria or the
 CC attenuated Pasteurellaceae bacteria can be used as vaccines in the
 CC fields of human medicine or veterinary medicine, and for identifying
 CC new antibacterial agents that target the virulence genes and their
 CC products. ABQ83458 to ABQ83578 and ABP54473 to ABP54551 represents
 CC sequences used in the exemplification of the present invention.
 CC
 XX
 SQ Sequence 175 AA;
 Query Match 8.1%; Score 164.5; DB 23; Length 175;
 Best Local Similarity 28.2%; Pred. No. 26-06; Mismatches 72; Indels 33; Gaps 4;
 Matches 53; Conservative 30; Mismatches 72; Indels 33; Gaps 4;
 Oy 60 TVELEGRLESLRTADVRAOVGGIIIOKRLFQEGSYVRAGOPLYQIDSSTYEANTLESARAO 119
 Db 5 TVIASGTIQATEQVDIGAVSGQIHIVQESQKVKKELLAVIDPRLAETELKIAKXEL 64
 Oy 120 ATAQA-----TLAKADADLARYKPLVAEAVEROEYDAAVTAKSAAGVKAQ----- 168
 Db 65 ANASANLDTKTKINLKQLQSDWERHRLRTATNATQKETEBAKSRNTAKAEIQINQND 124
 Oy 169 ---AAIKSAGINLNRSTRTAPISGFIGOSKVSSEGLLNAGDTTYLATIRQTNPMVNTQ 225
 Db 125 IAKIRVEKAETELGYTEIRSP-----DATVISVAQNGQTLIV-TTQ 165
 Oy 226 SASEVMKL 233
 Db 166 QVPVLMKL 173
 RESULT 28
 AAU36320
 ID AAU36320 standard; Protein; 285 AA.
 XX
 AC AAU36320;
 XX

DT 14-FEB-2002 (first entry)
 XX
 DE Pseudomonas aeruginosa cellular proliferation protein #310.
 XX
 DE Antisense; prokaryotic cellular proliferation protein;
 KW antibiotic; antibacterial; drug design.
 XX
 OS Pseudomonas aeruginosa.
 XX
 PN WO200170955-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 21-MAR-2001; 2001WO-US09180.
 XX
 PR 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-206848P.
 PR 26-MAY-2000; 2000US-207727P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 PI WPI, 2001-611495/70.
 DR N-PSDB; AAS54179.
 DR
 XX
 PT New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 XX
 PS Example 3; Seq ID No 11913; 511pp; English.
 XX
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acid can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 SQ Sequence 285 AA;
 Query Match 8.0%; Score 161.5; DB 22; Length 285;
 Best Local Similarity 26.0%; Pred. No. 7.3-06; Mismatches 101; Indels 55; Gaps 10;
 Matches 69; Conservative 40; Mismatches 101; Indels 55; Gaps 10;
 Oy 71 RTADVRA-----QVGGIIIOKRLFQEGSYVRAGOPLYQIDSSTYEANTLESARAO 123
 Db 35 RDAVRADVAVVADVSGWGTDLKQNVQKVDVLMRIDOERYQANLEBARVAEIRH 94
 Oy 124 ATLAKADADLARYKPL-----VAEAVEROEYDAAVTAKSAAGVKAQAAIKSAGINLN 179
 Db 95 QQVILRQNEAARSRRLIGAIISAEDKEMAQINMAI-----ARSEYQELAQVKIAEINLK 149
 Oy 180 RSRITAPISGFIGOSKVSSEGLLNAGDTTYLATIRQTN---PMVNVNTQASAEVMKLRRQ 236
 Db 150 RSEILRAARNGQVTNLRLAOGVYATAGQ-AVMALVDOOSFYVVAAYFEETKJLGIRVGMRAQ 208

KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
 KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
 KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
 KW hypertension; hypothyroidism; cholesterol ester storage disease;
 KW immune deficiency; immune disorder; infectious disease;
 KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
 KW myasthenia gravis.

OS Homo sapiens.

PN WO200192523-A2.

XX 06-DEC-2001.

XX 29-MAY-2001; 2001WO-US10836.

XX 30-MAY-2000; 2000US-206132P.

XX 29-AUG-2000; 2000US-228716P.

XX (CURA-) CURAGEN CORP.

XX Shinketsu RA, Leach MD;

XX WPI; 2002-106308/14.

XX N-PSDB; ABN20024.

XX Novel human polypeptides and polynucleotides useful for diagnosing,
 PT preventing and treating cardiovascular disease, neurodegenerative,
 PT hyperproliferative disorders and autoimmune disorders

PS Disclosure; SEQ ID 8526; 1037pp; English.

XX The present invention describes substantially purified human proteins
 CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
 CC in the specification). ABN15762 to ABN27252 encode the human ORFX
 CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
 CC treating or preventing a pathology associated with an ORFX-associated
 CC disorder in humans, and in the manufacture of a medicament for treating a
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
 CC sequences can be used in gene therapy. ORFX sequences can be used in the
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
 CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
 CC osteoarthritis, neurodegenerative disorders, disorders related to organ
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
 CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
 CC storage disease, various immune deficiencies and disorders, infectious
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
 CC bone degenerative disorders, or periodontal disease, and for gut
 CC protection or regeneration and treatment of lung or liver fibrosis,
 CC reperfusion injury in various tissues and conditions resulting from
 CC systemic cytokine damage.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 89 AA;

Query Match 8.2%; Score 165.5; DB 23; Length 89;

Best Local Similarity 43.0%; Pred. No. 66-07; Mismatches 1; Gaps 1;

Matches 34; Conservative 17; Indels 1; Gaps 1;

QY 211 ATIRQTNPYVNVTSASEVMKLRRQIAEGKLLADGVIAVGIKPDGTVPEKRLFLPA 270

DB 1 ASVQGLDPIYVDLTQSSNDFQLKQALIEGGS-LTKDGSQGVNIIMDTGQTYGHTKLEFP 59

QY 271 DPVNVSTGQITLRAAVPN 289

DB 60 DVTVDSTGTVTLRAVFPN 78

RESULT 26

ID AAB44578 standard; Protein; 175 AA.

AC AAB44578;

DT 08-FEB-2001 (first entry)

XX Virulence gene protein #58.

XX Virulence gene; antibacterial; vaccine; bacterial infection;
 KW septicemia; bronchopneumonia; rhinitis; wound infection.

XX Actinobacillus pleuropneumoniae.

XX WO200061724-A2.

XX 19-OCT-2000.

XX 06-APR-2000; 2000WO-US09218.

XX 09-APR-1999; 99US-0128689.

XX 10-SEP-1999; 99US-0153453.

XX (PHAA) PHARMACIA & UPJOHN INC.

XX Lowery DE, Fuller TE, Kennedy MJ;

XX WPI; 2000-647422/62.

XX N-PSDB; AAC79653.

XX Attenuated Pasteurellaceae bacteria comprising mutations in virulence
 PT genes, useful as a live attenuated vaccine against bacterial infections

PS Claim 39; Page 290; 322pp; English.

XX The family Pasteurellaceae encompasses several pathogens that infect a
 CC wide variety of animals. The present invention relates to virulence genes
 CC from Pasteurellaceae. The present sequence is a protein encoded by one
 CC such virulence gene. The virulence genes of the present invention may be
 CC mutated in order to produce an inactive gene. The inactive virulence gene
 CC may in turn be used to produce a vaccine, which is useful for treating
 CC bacterial infections such as septicemias, bronchopneumonias, rhinitis and
 CC wound infections.

XX Sequence 175 AA;

Query Match 8.1%; Score 164.5; DB 21; Length 175;

Best Local Similarity 28.2%; Pred. No. 2e-06; Mismatches 72; Indels 33; Gaps 4;

Matches 53; Conservative 30; Indels 33; Gaps 4;

QY 60 TVELPGRISRTADVRAQVGITIQKLPQEGSVYRAGQPLVQIDSTYENLESARQQL 119

DB 5 TVIASGTLQATEQVDIGAVSGQIKHIVQGGQYKKGKELLAVIDPRLAETELKAKAEL 64

QY 120 ATAAQ-----TLAKADADLARVKPLVAARAVSROEYDAVTKRBAAGVKAQ----- 168

DB 65 AMASANLDTKKINKKQLOSDWERHQRRLRNATQSKETEEKSRKLNTRKAKELQIAQNNND 124

QY 169 ---AIAKSGINLRNRSRTAPISGFIGOSKVSSEGLNAGDTYVATIRQTNPMYVNVTO 225

DB 125 IAKRIVEKAELELQVTEIRSPLE-----DATVSVPAQNGQTLV-TTQ 165

QY 226 SASSEVMKL 233

DB 166 QVPVLMKLL 173

RESULT 27

ABP54530 standard; Protein; 175 AA.

CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

CC Sequence 299 AA;

Query Match 8.4%; Score 169; DB 22; Length 299;
 Best Local Similarity 25.4%; Pred. No. 1.7e-06;
 Matches 71; Conservative 43; Mismatches 105; Indels 60; Gaps 9;

QY 52 VHPQTVALTVLPGRLSRLRTADVRAQVGGIIQKRLFQEGSYVRAQGPLYQIDSSSTEYAN 111

DB 54 IRAEVSSTPQVSGRIVELINMD-----NQVWNGDILLITDKTPFOA 97

QY 112 LESAPAQLATAQATLAKADADIALARKPLVAEAVSRQEVDAAVTKRSAGVKAQAQAI 171

DB 98 ELNAAQALAKQSDIAKNNENRRRL-SQNFISAEELDTANLVKMAQASVDAQAQTL 156

QY 172 KSAGINLRSRTAPISGRIGSKVSEGLNAGPTTVATIRQTNPMVAVNTQASAEVM 231

DB 157 KOAQMLQTEIRAPVSGWVNLTRIGYADTG-----KPLFALVDSISFYI 205

QY 232 -----KLRFQAEK-----LTAADG-----VIAVGIKFDGTVYPEKGRLLFADPVN 275

DB 206 GFEEETKL-RHIREGAPQITLYSDNKTLCQHVSSIGRAIYQSVESDSSL-----PDVK 260

QY 276 ESTGQITLRAAVPNDONTIMPELVYRVLMDOYAVDNAPV 314

DB 261 PNVPWVRLAQRPV-----VRFALDKVPGDITLV 288

RESULT 24

ID AUJ38257 standard; Protein; 309 AA.

AC AUJ38257;

DT 14-FEB-2002 (first entry)

DE Salmomella typhi cellular proliferation protein #148.

KW Antisense; prokaryotic cellular proliferation protein;

OS Salmomella typhi.

PN MO200170955-A2.

PD 27-SEP-2001.

PF 21-MAR-2001; 2001WO-US09180.

PR 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207727P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

XX Haselbeck R, Ohlsen KU, Zyskind JW, Wall D, Trawick JD, Carr GJ;

PI Yamamoto RT, Xu HH;

XX WPI; 2001-611495/70.

DR N-PsDB; AAS56116.

PT New polynucleotides for the identification and development of

XX antibiotics, comprise sequences of antisense nucleic acids -

XX Example 3; Seq ID No 13850; 511pp; English.

CC The invention relates to antisense inhibitors of genes essential to

CC prokaryotic cellular proliferation, their use in identifying the

CC genes, their use in the discovery of novel antibiotics, the essential

CC genes themselves and the encoded proteins. The prokaryotes used are

CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*

CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The

CC invention is also useful for the identification of potential new targets

CC for antibiotic development. The antisense nucleic acids can also be used

CC to identify proteins used in proliferation, to express these proteins,

CC and to obtain antibodies capable of binding to the expressed proteins.

CC The proteins can be used to screen compounds in rational drug discovery

CC programmes. The antisense nucleic acid sequence is also useful to screen

CC for homologous nucleic acids which are required for cell proliferation in

CC a wide variety of organisms. The present sequence represents an

CC essential prokaryotic cellular proliferation protein.

CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic

CC format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 309 AA;

Query Match 8.4%; Score 169; DB 22; Length 309;

Best Local Similarity 24.5%; Pred. No. 1.8e-06;

Matches 63; Conservative 45; Mismatches 103; Indels 46; Gaps 7;

QY 75 VRAQVGGITQRLFOEGSYVRAQGPLYQIDSSSTEYANESRAQATLQATLAKA--DAD 132

DB 70 VTPQVSGSITQINKDQFNVAGDVLFPVIDTPHIELNQAQDLAKQOSDLAANNEAD 129

QY 133 LARXKPLVAEAVSRQEVDAAVTKRSAGVKAQAQAIKSGAGINLRSRTAPISGRIG 192

DB 130 RRRH---LSRNYIAEDDSDANLVKMAQASVDVALTLKQAQMLQSTEVKAPVSGWVT 186

QY 193 QSKVSEGLNAGDPTTVLATIRQTNPMVAVNTQASAEVM-----KLRR-----QI 237

DB 187 NLSTRTGYASTG-----KPLFALVDSISFYMGVFEETKLRHIREGPAITLL 235

QY 238 AEGKILAADGVIAVGIKFDGTVYPEKGRLLFADPVNVESTGQITLRAAVPNDONTIMPG 297

DB 236 YSGWVKLQGHGSGIRAIYDQSVESDGLV-----PDIKPNPWRRLAQRPV----- 282

QY 298 LYRVLMDOYAVDNAPV 314

DB 283 --VRIEFDALPDITLV 297

RESULT 25

ABP04272

ID ABP04272 standard; Protein; 89 AA.

AC ABP04272;

DT 24-JUN-2002 (first entry)

DE Human ORFX protein sequence SEQ ID NO:8526.

XX Human, open reading frame; ORFX; gene therapy; cancer; cirrhosis;

QY 254 KPDGTVYPEKGRLLFADPVNVESTGQITLRAAVNDQNIIMPLGYRVLMDQVADNAF 313
 DB 299 GFSAGT-----GSAASLIPDAQNATGNWIKVORVP-----VAIVLRDVD--- 339
 QY 314 VPPQAVTRGAKDTVMIVNAOGGMEPRE---VTAQOOGTNW 352
 DB 340 ---RHPLRIGLSMTYKVDTSAAAGAVSKTPGALPEMESTDW 378

RESULT 22
 AAG3061
 ID AAG3061 standard; Protein; 659 AA.
 AC AAG3061;
 DT 26-SEP-2001 (first entry)
 DE C glutamicum protein fragment SEQ ID NO: 6815.
 XX
 KM Corynebacterium; amino acid synthesis; vitamin; saccharide;
 XX organic acid synthesis.
 XX OS Corynebacterium glutamicum.
 XX
 PM EPI108790-A2.
 PD 20-JUN-2001.
 XX 18-DEC-2000; 2000EP-0127688.
 XX
 PR 16-DEC-1999; 99JP-0377484.
 PR 07-APR-2000; 2000JP-0159162.
 PR 03-AUG-2000; 2000JP-0280988.
 XX
 PA (KYOW) KYOWA HAKKO KOGYO KK.
 XX
 PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 DR WPI; 2001-376931/40.
 DR N-PSDB; AAH68280.
 XX
 PT Novel polynucleotides derived from Corynebacterium, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 PT
 PS
 XX
 CC Claim 17; SEQ ID NO: 6815; 246bp + Sequence Listing; English.
 CC
 CC The present invention provides a number of nucleotide and protein
 CC sequences from the Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of corynebacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from Corynebacterium, and identifying a homologue of a gene derived
 CC from corynebacterium. Corynebacterium are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a protein described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 CC
 XX
 SQ Sequence 659 AA;
 XX
 Query Match 8.7%; Score 175; DB 22; Length 659;
 Best Local Similarity 22.1%; Pred. No. 1.6e-06;
 Matches 102; Conservative 81; Mismatches 194; Indels 84; Gaps 16;
 QY 5 AFKAMRAAALAAVALVIVSSCGKGDAAAGGQAPAGREAPAVGVVTHPQTVALTVEL- 63
 DB 223 ALKAREQADAEIERLRADPGLNN-----DRSNINDVIGLDERESIASASEELA 274

QY 64 ----PGRLESRTADVAVOAGGIIQKRLFOEGSYVRAGQPIYOIDSSTYEANLESARAQL 119
 DB 275 QARAGDLEAVAAAE--AKVAGLEQSIASKSTSTPSPDQTYLQ---SYTLEAEARRVA 328
 QY 120 ATAQA-----TLAKADADLARYKPLVAEAVSROEYDAV---TAKSAEAGYKA 166
 DB 329 STTEALEIERIYIDSIGKVDSELAARAV-ABEHSAQ-DAALGLETALQSTOHOLEA 386
 QY 167 AQAAIKSA-----GININRSRITAPISGFI-----GQSKVEGTLN 203
 DB 387 QSSAIDALGLAVDNEAATRSTSQLRMDINTTVRSPPSGIVSSVOAAQOGAPAGALLS 446
 QY 204 AGDTT---VLATIRQTPMYVNV-----TQASEWMLRROIAGKLLADGVAVGIK 254
 DB 447 VADSELEKITHANVEAISNVLTIGSRFTFTTPTSGTGFAGRVSKVPIAA----- 497
 QY 255 FDDGTVYPEKGRLLFADPVNVESTGQITLRAAVNDQNIIMPLGYRVLMDQVADNAFV 314
 DB 498 ---AASAPATGEGAAAGATTNTDVTPEIISVTGDEGLNLGGSARVRIVHEIAPHVLT 554
 QY 315 VPPQAVTRG--AKDTVMIVNAOGGMEPREVTVAQOOGTNWIVT--SGLKDGKRVVEGISI 371
 DB 555 VPLEAVYKNDGKDAVLIISDNEKVEEVEKTAESDDFDIAVSGAGISEDARVILTPGNY 614
 QY 372 AGITGAKKVTPEKWAASENQAAPQSGVQTASEAKTASEAE 412
 DB 615 RGLIGE---TVKLHADTVEQAAAPFSPAAPPDPAAPVSAK 652

RESULT 23
 AAU34575
 ID AAU34575 standard; Protein; 299 AA.
 AC AAU34575;
 DT 14-FEB-2002 (first entry)
 DE E. coli cellular proliferation protein #156.
 XX
 KM Antisense; prokaryotic cellular proliferation protein;
 KM antibiotic; antibacterial; drug design.
 XX
 OS Escherichia coli.
 XX
 PM WO200170955-A2.
 PM
 PD 27-SEP-2001.
 PF 21-MAR-2001; 2001WO-US09180.
 XX
 PR 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-206848P.
 PR 26-MAY-2000; 2000US-207727P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 DR WPI; 2001-611495/70.
 DR N-PSDB; AAS52434.
 PT New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 XX Example 3; Seq ID No 10168; 51bp; English.
 PS
 XX The invention relates to antisense inhibitors of genes essential to

KW cardiovascular disease; immune system disorder; organ transplantation;
KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
KW hypothyroidism; cholesterol ester storage disease; infection; vulvovaginitis;
KW vasoprotective; antipsoriatic; antidiabetic; cytosolic; neurotrophic;
KW neuroprotective; antithrombotic; anticoagulant; thrombolytic;
KW cardiac; hypotensive; antihypertensive; antidiabetic; immunomodulator;
KW dermatological; analgesic; virucide; antibacterial; fungicide.
OS Homo sapiens.
XX WO200190366-A2.
XX 29-NOV-2001.
XX 24-MAY-2001; 2001WO-US17076.
XX 24-MAY-2000; 2000US-206690P.
XX (CURA-) CURAGEN CORP.
XX Leach MD, Shimkets RA;
XX MPI; 2002-106200/14.
XX N-PSDB; ABN75407.
XX Novel human polypeptides and polynucleotides useful for diagnosing,
XX preventing and treating cardiovascular disease, neurodegenerative,
XX hyperproliferative disorders and disorders related to organ
XX transplantation -
XX Claim 10; Page 445; 2508pp; English.
XX
XX Sequences ABP1028-ABP35561 represent 4534 novel human proteins
XX designated ORF (open reading frame) 1-4534, and sequences ABN75054-
XX ABN79587 represent cDNAs encoding them. The invention also encompasses
XX polypeptides at least 80% identical to the ORF1-ORF4534 (collectively
XX referred to as ORFX) proteins, polynucleotides at least 85% identical to
XX the ORFX nucleic acid sequences, vectors and host cells comprising ORFX
XX polynucleotides, the recombinant production of ORFX proteins, antibodies
XX specific for ORFX proteins, methods of detecting ORFX polynucleotides and
XX polypeptides, methods of screening for modulators of ORFX expression or
XX activity, and methods of screening individuals for a predisposition to an
XX ORFX-associated disorder. The ORFX proteins of the invention have a wide
XX range of biological activities, such as cytokine, cell proliferation,
XX cell differentiation, immune modulation, haematopoiesis regulation,
XX tissue growth, angiogenesis, activin or inhibin activity, chemotactic/
XX chemokinetic activity, haemostatic activity, thrombolytic activity,
XX receptor/ligand, antiinflammatory activity, tumour inhibition activity,
XX and antiinfective activity, and may also be involved in the determination
XX of bodily characteristics, fertility and behaviour. ORFX proteins,
XX nucleic acids and antibodies may be used in the treatment of cancers,
XX other proliferative disorders such as psoriasis and benign tumours,
XX neurological disorders such as epilepsy and Alzheimer's disease,
XX cardiovascular diseases, immune system disorders, disorders related to
XX organ transplantation, disorders of tissue growth and regeneration,
XX diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester
XX storage disease, and infectious diseases caused by viral, bacterial,
XX fungal and other pathogens. ORFX nucleic acids may also be used as a
XX source of primers and probes, in the detection of ORFX genomic sequences
XX or transcripts, in the identification and cloning of homologous
XX sequences, in genetic diagnosis, and in forensic biology. The ORFX
XX nucleic acids may additionally be used to produce transgenic animals
XX which may be useful for studying the function and/or activity of ORFX
XX protein, and in drug screening. The ORFX proteins may also be used as
XX immunogens to generate specific antibodies, which are useful in the
XX diagnosis, treatment and monitoring of ORFX-associated diseases.
XX
XX Sequence 102 AA;
SQ

Query Match 10.7%; Score 217; DB 23; Length 102;
Best Local Similarity 40.2%; Pred. No. 2.1e-11;
Matches 41; Conservative 28; Mismatches 33; Indels 0; Gaps 0;

QY 221 VNVTSASEWKLKRLQIAEGKLLAADGVIANVGIKFDGTVPEKRLFPADPVNVESTGQ 280
Db 1 VDVTSNDPFLKQELSGALKOENGKAKVLLLENQAOYAQETLEFSDVTDETGS 60
QY 281 ITRAAVNDONILMPGLYRVLVMDQVAVDPAFVPOQAVTR 322
Db 61 ITRAVFPNPDTLPGFVFAFARLDEGVRSALLVPOGAVTR 102

RESULT 21
ABP79084
ID. ABP79084 standard; Protein, 394 AA.
XX
XX AC ABP79084;
XX DT 07-MAR-2003 (first entry)
XX DE N. gonorrhoeae amino acid sequence SRQ ID 4698.
XX KW Antibacterial; infection; vaccine; gene therapy.
XX OS Neisseria gonorrhoeae.
XX PN WO200279243-A2.
XX PD 10-OCT-2002.
XX PF 12-FEB-2002; 2002WO-1B02069.
XX PR 12-FEB-2001; 2001GB-0003424.
XX PA (CHIR-) CHIRON SPA.
XX PI Fontana MR, Pizsa M, Masignani V, Monaci E;
XX WPI; 2003-058415/05.
XX DR N-PSDB; ABZ40054.
XX PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a
XX medicament for treating or preventing N. gonorrhoeae infection -
XX PS Disclosure; Page 533; 815pp; English.

XX
XX The present invention relates to proteins from Neisseria gonorrhoeae.
XX Also disclosed are the nucleic acid molecules encoding the proteins and
XX antibodies that specifically bind to the proteins. The composition
XX comprising the protein, nucleic acid or antibody is useful for the
XX manufacture of a medicament for treating or preventing N. gonorrhoeae
XX infection, this may be in the form of a vaccine or gene therapy.
XX Sequences given in records ABP76736-ABP81046 represent nucleic acid
XX molecules of the invention.

SQ Sequence 394 AA;
Query Match 9.3%; Score 188; DB 24; Length 394;
Best Local Similarity 24.3%; Pred. No. 5.3e-08;
Matches 83; Conservative 42; Mismatches 127; Indels 90; Gaps 10;

QY 71 RTAVRQVGGITIKRLPQESGYRAGOPLYQIDSS---TYE----- 109
Db 67 RVVQVTPQKGGTVKRVKLVHDDTDVAVKGDVLAIVDDNDVLAVERAKNELVOAVRNRON 126
QY 110 ANLESARQQLATAQATLAKADADLARYKPLVAEAVSROEYDAVTFARSAEAGVKAQA 169
Db 127 AATSOAGQVALRADARADDDLRSSALAEAGVAEELAHARTAVSOQAQAAVKAALA 186
QY 170 -----AIKSGAINLNSRITAPISGFKVSEGTLLN 203
Db 187 EESSARAALGGDVSLRQPEVQTAIGRLKQAMINLRTQVRAPADGVAKRSVQGGQVA 246
QY 204 AG-----DTTVALATRTQNPMPVNVVTSASEWKLKRLQIAEGKLLAADGVIANVGI 253
Db 247 AGAELMAVPLSDVWVADANFKEITLRHMKIGOPAEIVSDI-----YKQIYVRGRVA--- 298

CC and N. gonorrhoeae antigenic proteins. They are encoded by open
 CC reading frames (ORFs) AA211972-212358. The antigenic proteins,
 CC their fragments, their nucleic acids and antibodies are used for
 CC diagnosis, prevention (as vaccines) or treatment of Neisseria
 CC infections, such as meningitis, septicaemia and gonorrhoea. Both
 CC organisms are closely related. Fragments of the nucleic acids
 CC are useful as hybridisation probes and antisense reagents.

XX Sequence 334 AA;

Query Match 13.4%; Score 271; DB 20; Length 334;
 Best Local Similarity 28.6%; Pred. No. 1.9e-15;
 Matches 94; Conservative 56; Mismatches 135; Indels 44; Gaps 9;

QY VRAQVGIIORLRFEGSVYRAGQPLVQIDS-----STVEANLESARQATATLA 127
 DB 3 VQAQASGQIKILVYKLGQOVKKGDILAEINSTQNTLNTEKSKLETYQAKLVSAQIALG 62
 QY 128 KADADLARYKPLVAEAVSROEYDAVTAKRSAEAGVKAQAQAIKSAGINLNR----- 180
 DB 63 SAEKKYKROALMKENATSKEDLESADAFAPAAKANVALEKALIRKSKISINTASELGY 122
 QY 181 SRTIAPISGFIQSGKVSSEGLTLNAGDTTVLATIRQTNPMYVNTQSASEVKKLRQIAEG 240
 DB 123 TRITATMDGTVAALIVEEGQTVNAA-----QSTPTIYQLAN--LDMMLNKKQIAEG 171
 QY 241 KL--LAADGVAVGIRKPDGTVYPEKGRLEFADPVV-----NESTGQIT-----LR 284
 DB 172 DITKRAQODISFTILSEPT--PIKAKLDSDVDEGLTTMSSGGYNSSTDITASNAVYYAR 229
 QY 285 AAVPNDQNIIMPGLYVRLMDQVAVDNAFVVPQOAV--TRGAKDTVMIVNAOGMEPREVT 343
 DB 230 SFVNPDPGLKATGMTQNTQVTEIDGKAVLLIPSLTVNKRGGKAFVRLVAGDGKKAERIR 289
 QY 344 VAAQOGTNMIVTSGLKDGDKRVVVGISIA 372
 DB 290 TGMRDSMNTVEVSGIKEGDKVIVSEITAA 318

RESULT 19

AAV74870 ID AAV74870 standard; Protein; 334 AA.

XX AC AAV74870;

XX DT 21-MAR-2000 (first entry)

XX DB Neisseria meningitidis ORF 290 protein sequence SEQ ID NO:1214.

XX KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
 KW antibacterial; gene therapy.

XX OS Neisseria meningitidis.

XX PM MO9957280-A2.

XX PD 11-NOV-1999.

XX PF 30-APR-1999; 99WO-US09346.

XX PR 01-MAY-1998; 98US-0083758.

XX PR 31-JUL-1998; 98US-0094869.

XX PR 02-SEP-1998; 98US-0098994.

XX PR 02-SEP-1998; 98US-0099062.

XX PR 09-OCT-1998; 98US-0103749.

XX PR 09-OCT-1998; 98US-0103794.

XX PR 25-FEB-1999; 99US-0121528.

XX PA (CHIR) CHIRON CORP.

XX PA (GENO-) INST GENOMIC RES.

PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Piza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tetteili H, Venter JC;
 XX WPI; 2000-062150/05.
 DR N-PSDB; AA253632.
 XX
 PT Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics
 XX
 PS Claim 2; Page 677; 1453p; English.
 XX
 CC AA253015 to AA254536, AA254577 to AA254615, and AAV74253 to AAV75941
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.

XX Sequence 334 AA;

Query Match 13.4%; Score 271; DB 21; Length 334;
 Best Local Similarity 28.6%; Pred. No. 1.9e-15;
 Matches 94; Conservative 56; Mismatches 135; Indels 44; Gaps 9;

QY VRAQVGIIORLRFEGSVYRAGQPLVQIDS-----STVEANLESARQATATLA 127
 DB 3 VQAQASGQIKILVYKLGQOVKKGDILAEINSTQNTLNTEKSKLETYQAKLVSAQIALG 62
 QY 128 KADADLARYKPLVAEAVSROEYDAVTAKRSAEAGVKAQAQAIKSAGINLNR----- 180
 DB 63 SAEKKYKROALMKENATSKEDLESADAFAPAAKANVALEKALIRKSKISINTASELGY 122
 QY 181 SRTIAPISGFIQSGKVSSEGLTLNAGDTTVLATIRQTNPMYVNTQSASEVKKLRQIAEG 240
 DB 123 TRITATMDGTVAALIVEEGQTVNAA-----QSTPTIYQLAN--LDMMLNKKQIAEG 171
 QY 241 KL--LAADGVAVGIRKPDGTVYPEKGRLEFADPVV-----NESTGQIT-----LR 284
 DB 172 DITKRAQODISFTILSEPT--PIKAKLDSDVDEGLTTMSSGGYNSSTDITASNAVYYAR 229
 QY 285 AAVPNDQNIIMPGLYVRLMDQVAVDNAFVVPQOAV--TRGAKDTVMIVNAOGMEPREVT 343
 DB 230 SFVNPDPGLKATGMTQNTQVTEIDGKAVLLIPSLTVNKRGGKAFVRLVAGDGKKAERIR 289
 QY 344 VAAQOGTNMIVTSGLKDGDKRVVVGISIA 372
 DB 290 TGMRDSMNTVEVSGIKEGDKVIVSEITAA 318

RESULT 20

ABP31381 ID ABP31381 standard; Protein; 102 AA.

XX AC ABP31381;

XX DT 08-JUL-2002 (first entry)

XX DE Human glycoprotein-like ORF354 protein, SEQ ID NO:708.

XX KW Human; ORF; open reading frame; ORF; drug screening; diagnosis;
 KW disease monitoring; cytokine; cell proliferation; cell differentiation;
 KW immune modulation; haematopoiesis regulation; tissue growth;
 KW angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;
 KW thrombolytic; tumour inhibition; bodily characteristic; fertility;
 KW behaviour; cancer; proliferative disorder; neurological disorder;

QY 120 ATAQATLAKADADLARYKPLVAEAVSROEYDAVTAKRSAAGYKAAQAIAKSAGINLN 179
 Db 113 VSAQIALGSAEKYKRAQALMKDADATAKEDLESADALAAKAVNAELKALIROSKISIN 172
 QY 180 R-----SRTAPISGFIGSKVSEGLTNAGDITVATIRQINPMYVNTQSASEVWK 232
 Db 173 TAESLGTTRITATMDGTVALVLEEGQTVNAA-----QSPFTIVQLAN--LDMWL 221
 QY 233 LRQIAEGL--LAADGVIAVGIRKFDGTVYPEKGRLLPADPV-----NESTGOIT 282
 Db 222 NKMQIAEGDITKVKAGODISFTILSEPT--PIKAKLDSVDPGLTMSGGYNSSTDAS 279
 QY 283 -----LRAAVPNDQNIIMPGLYVRVLDQVAVDNAPFVPPQAV--TRGAKDTVMVNAOG 335
 Db 280 NAVYYAASFPVNPDPGKATGTTONTVEIDGKXVLIIPSLTVKNRGGRAFVRVLGADG 339
 QY 336 GMEPEVTVAAOQGTNMIYVTSGLKXGDKVVEGISA 372
 Db 340 KAAEREIRTGMRDMSMTTEVKSGLKEGDKVVISSEITAA 376

RESULT 17

AAU73013
 ID AAU73013 standard; Protein; 392 AA.

AC AAU73013;

DT 12-MAR-2002 (first entry)

DE Neisseria meningitidis virulence protein #103.

KM Meningitis; virulence; gene; antibacterial; vaccine; veterinary;

KX infection; Gram-negative bacteria; antimicrobial.

OS Neisseria meningitidis.

FN WO200185772-A2.

PD 15-NOV-2001.

PF 08-MAY-2001; 2001WO-GB02003.

PR 08-MAY-2000; 2000GB-0011108.

PA (MICR-) MICROSCIENCE LTD.

PI Tang C;

DR WPI; 2002-066593/09.

DR N-PSDB; AAS97298.

PT New peptide encoded by operon including virulence genes of Neisseria meningitidis, useful as vaccine component for treating or preventing meningitis and for identifying antimicrobial drug

PS Claim 4; Page 419-420; 423pp; English.

XX The invention relates to a peptide (I) encoded by an operon (II) of Neisseria meningitidis including virulence genes, or a related molecule having a 40% sequence similarity at the peptide or nucleotide level in a Gram-negative bacterium, or its functional fragment, for therapeutic or diagnostic use. (I) and (II) are useful in the manufacture of a medicament for treating or preventing a condition (e.g., meningitis) associated with infection by Neisseria or Gram-negative bacteria. The product is useful for veterinary treatment and in a screening assay for the identification of an antimicrobial drug. The vaccines have prophylactic applications. AAU72911-AAU73014 represent N. meningitidis virulence proteins of the invention.

SQ Sequence 392 AA;

Query Match 15.3%; Score 308; DB 23; Length 392;
 Best Local Similarity 27.5%; Pred. No. 1.3e-18;

Matches 109; Conservative 65; Mismatches 169; Indels 54; Gaps 10;
 QY 7 KAMRAALAAVALVLSGCGKGDAAOGGAPAGRAPVGVTVHPQVALTVLELPCR 66
 Db 3 KMMKWAALVAAVAAA-----AVWGWGSLYKREPOAAVITETVRGRDISRTVSATGE 52
 QY 67 LESLRTADVRAQVGGIIQKRLFOEGSYVRAQOPLYQIDS-----STYEANLESARQOL 119
 Db 53 ISPSNLVSGAQSQIKLVKLGQOVKRDLLAEINSTQNTLTNTEKSKLETYQAKL 112
 QY 120 ATAQATLAKADADLARYKPLVAEAVSROEYDAVTAKRSAAGYKAAQAIAKSAGINLN 179
 Db 113 VSAQIALGSAEKYKRAQALMKDADATAKEDLESQADALAAKAVNAELKALIROSKISIN 172
 QY 180 R-----SRTAPISGFIGSKVSEGLTNAGDITVATIRQINPMYVNTQSASEVWK 232
 Db 173 TAESLGTTRITATMDGTVALVLEEGQTVNAA-----QSPFTIVQLAN--LDMWL 221
 QY 233 LRQIAEGL--LAADGVIAVGIRKFDGTVYPEKGRLLPADPV-----NESTGOIT 282
 Db 222 NKMQIAEGDITKVKAGODISFTILSEPT--PIKAKLDSVDPGLTMSGGYNSSTDAS 279
 QY 283 -----LRAAVPNDQNIIMPGLYVRVLDQVAVDNAPFVPPQAV--TRGAKDTVMVNAOG 335
 Db 280 NAVYYAASFPVNPDPGKATGTTONTVEIDGKXVLIIPSLTVKNRGGRAFVRVLGADG 339
 QY 336 GMEPEVTVAAOQGTNMIYVTSGLKXGDKVVEGISA 372
 Db 340 KAAEREIRTGMRDMSMTTEVKSGLKEGDKVVISSEITAA 376

RESULT 18

AAV38882
 ID AAV38882 standard; Protein; 334 AA.

AC AAV38882;

DT 08-OCT-1999 (first entry)

DE Neisseria meningitidis antigen encoded by a partial ORF85.

KM Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;

KX treatment; Neisseria infection; meningitis; septicemia; gonorrhea.

OS Neisseria meningitidis.

FN WO9924578-A2.

PD 20-MAY-1999.

PF 09-OCT-1998; 98WO-IB01665.

PR 01-SEP-1998; 98GB-0019016.

PR 06-NOV-1997; 97GB-0023516.

PR 14-NOV-1997; 97GB-0024190.

PR 18-NOV-1997; 97GB-0024386.

PR 27-NOV-1997; 97GB-0025158.

PR 10-DEC-1997; 97GB-0026147.

PR 14-JAN-1998; 98GB-0000759.

PA (CHIR-) CHIRON SPA.

PI Grandi G, Mesigiani V, Pizza M, Rappuoli R, Scarlato V;

DR WPI; 1999-327407/27.

DR N-PSDB; AA212304.

PT Proteins from Neisseria meningitidis and N. gonorrhoeae useful for diagnosis, treatment and prevention of infection

PS Claim 4; Page 427; 524pp; English.

SQ Amino acid sequences AAV38499-Y38944 represent Neisseria meningitidis

XX	antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW	antibacterial; gene therapy.
XX	Neisseria meningitidis.
OS	
PN	W0957280-A2.
XX	
PD	11-NOV-1999.
XX	
PF	30+APR-1999; 99WO-US09346.
XX	
PR	01-MAY-1998; 98US-0083758.
XX	
PR	31-JUL-1998; 98US-0094869.
PR	02-SEP-1998; 98US-0098994.
PR	02-SEP-1998; 98US-0099062.
PR	09-OCT-1998; 98US-0103749.
PR	09-OCT-1998; 98US-0103794.
PR	09-OCT-1998; 98US-0103796.
PR	25-FEB-1999; 99US-0121528.
XX	
PA	(CHIR) CHIRON CORP.
PA	(GENO-) INST GENOMIC RES.
XX	
PI	Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI	Petersen J, Pizzi M, Rapunoli R, Ratti G, Scalato E, Scarselli M;
PI	Tetelji H, Venter JC,
XX	
DR	WPI: 2000-062150/05.
DR	N-PSDB; AAZ53633.
XX	
PT	Novel Neisserial polypeptides predicted to be useful antigens for
PT	vaccines and diagnostics -
PS	Claim 2; Page 678; 1453pp; English.
XX	
CC	AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC	represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
CC	and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC	PCR primers used in the exemplification of the present invention. The
CC	polypeptides, the polynucleotides, antibodies and compositions of
CC	the invention can be used as vaccines, as diagnostic reagents, and as
CC	immunologic compositions. The polypeptides can be used in the
CC	manufacture of medicaments for treating or preventing infection due to
CC	Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC	presence of Neisseria bacteria, or to raise antibodies. They may also
CC	be used to screen for agonists or antagonists, which may themselves
CC	have use as antibacterial agents. The polynucleotides of the invention
CC	may also be used in gene therapy protocols.
XX	
XX	Sequence 392 AA;
XX	
XX	Query Match 15.3%; Score 308; DB 21; Length 392;
XX	Best Local Similarity 27.5%; Pred. No. 1.3e-18;
XX	Matches 109; Conservative 65; Mismatches 169; Indels 54; Gaps 10;
QY	7 KAMRAAALAAVAALVLSGCGKGDAAOGGAPAGRAPAVGVVTHPQTALTYELDER 66
DB	3 KMMKMAAAVAALVLSGCGKGDAAOGGAPAGRAPAVGVVTHPQTALTYELDER 66
QY	67 LESTALTAIVRAQVGGIIIOKRLFOEGSSVYRAGOEPLYOIDS-----STYEANLJESARAO 119
DB	53 ISPSNLVSVGAOASQIKLYKLGQCKKGLILAEINSTQNTLNTEKSLTYEQAKL 112
QY	120 ATAQNTAKADADILARYKPLVAAEAVSROEYDAAVTAKRSAGYKAAQAIAIKSGINLN 179
DB	113 USAQIALTSAEKKYKRAQALMWDDTAKEDEESADALAAKANVAELKALIRQSKISIN 172
QY	180 R-----SRTAPISGFIQGSKSEGTLLNAGDTTVALTITQTPMNVNVQASSEYWK 232
DB	173 TAESLGYTRITATMDGTVALIVEEGQTVNA-----QSTPTTVOLAN--LDMWL 221
QY	233 LRRQIAEKL--LAADGVIAVGIRKPDGTVPEKGRLLFADEVV-----NESTGQIT 282

```

Db      222  NKMQAEBDITRKVKAGQDISFTIILSEPDT--PIKAKLSVDVDELITMSSGYNSSSTDFTAS 279
Qy      283  -----LFAAVPNDONILMPGIYRVLMDQVAVDNPAFVPOAV-TRGAKDTVMYVNAAG 335
Db      280  NAVVYVYASFPVNPEDGKLATGTTQNTVEIDGVKNVLIIPSLTVKNRGGRAFRVVLGADG 339
Qy      336  GMEPREVTVAOQGGTNMIVTSGLKDGDVKVVEGISTIA 372
Db      340  KAEREIRITGMRDSMTTEVKSGLNEGDKVYSEITAA 376

RESULT 16
AAU72963
AAU72963 standard; Protein; 392 AA.
AC      AAU72963;
AD      12-MAR-2002 (first entry)
DE      Neisseria meningitidis virulence protein #53.
KM      Meningitis; virulence; gene; antibacterial; vaccine; veterinary;
        infection; Gram-negative bacteria; antimicrobial.
OS      Neisseria meningitidis.
PN      WO200185772-A2.
PD      15-NOV-2001.
PF      08-MAY-2001; 2001WO-GB02003.
PR      08-MAY-2000; 2000GB-0011108.
PA      (MICR-) MICROSCIENCE LTD.
PI      Tang C;
PL      WPI; 2002-066593/09.
DR      N-PSDB; AAS97248.
XX      New peptide encoded by operon including virulence genes of Neisseria
PT      meningitidis, useful as vaccine component for treating or preventing
PT      meningitis and for identifying antimicrobial drug
PS      Claim 4; Page 226-228; 423pp; English.
XX      The invention relates to a peptide (I) encoded by an operon (II) of
CC      Neisseria meningitidis including virulence genes, or a related molecule
CC      having a 40% sequence similarity at the peptide or nucleotide level in a
CC      Gram-negative bacterium, or its functional fragment, for therapeutic or
CC      diagnostic use. (I) and (II) are useful in the manufacture of a
CC      medicament for treating or preventing a condition (e.g., meningitis)
CC      associated with infection by Neisseria or Gram-negative bacteria. The
CC      product is useful for veterinary treatment and in a screening assay for
CC      the identification of an antimicrobial drug. The vaccines have
CC      prophylactic applications. AAU72911-AAU73014 represent N. meningitidis
CC      virulence proteins of the invention.
SQ      Sequence 392 AA.

Query Match 15.3%; Score 308; DB 23; Length 392;
Best Local Similarity 27.5%; Pred. No. 1,3e-18;
Matches 109; Conservative 65; Mismatches 169; Indels 54; Gaps 10

Db      7  KAMRAAALAAVAVALVSSCGKGDAAQGGPAGREAPAVGVVTHPQTVALTVELPGR 66
Qy      3  KKKKAAVAVAVAAA-----AVWGGSWYLKEPQAAVITETVRRGDISRTVSATGE 52
Db      67  LESLTLADYRAOVGGITQKRLFOEGSYRAAGPPIYDID-----STEANIESARAOL 119
Qy      53  ISPSNLVSVGAQASQIKKLTVKLGQQQVKGGLIAEINSTQNTLNTFKSKLETYQAKL 112

```

CC diagnosis, prevention (as vaccines) or treatment of *Neisseria*
 CC infections, such as meningitis, septicaemia and gonorrhoea. Both
 CC organisms are closely related. Fragments of the nucleic acids
 CC are useful as hybridisation probes and antisense reagents.

XX Sequence 392 AA;

Query Match 15.3%; Score 308; DB 20; Length 392;

Best Local Similarity 27.7%; Pred. No. 1.3e-18;
 Matches 110; Conservative 63; Mismatches 170; Indels 54; Gaps 10;

```

OY 7 KAMRAAALAAVALVLSGCKGDAAGGAPAGREAPAPVGVTVHPQVALTYELPCR 66
DB 3 KMKKAAVAALVAAA-----AWGGMWYLYKPEPOAAVITTAVRGDISRTVSATGE 52
OY 67 LESLRTADVRAQVGIIQKRLFOEGSYVRAGOPLYOIDSSTY-----EANLESARAOL 119
DB 53 ISPSNLVSVGAQASGQIKLVKLGQGVKKGDLAEINSTTQNTIDMEKSKLETYQAKL 112
OY 120 ATAQATLAKADADLARYKPLVAEAVSRQEVDAAVTKRSAGAKVAAQAIAKSGINLN 179
DB 113 VSAQIALGSAEKKYKROAALMKDDATSKEDLESADALAAKAVAAELKALIRQKISIN 172
OY 180 R-----SRTAPISGFIGOSKVSSEGLTNAAGDTVLATITROTNPVYVNTQSASEVWK 232
DB 173 TAESDLGYTRITATMDGTVVAIPVEEGQTVNAA-----QSTPTTVQLAN--LDMWL 221
OY 233 LRQIAEGKL--LAADGVAVGIRKFPDGTVPYPERGRLLFADPVV-----NESTGOIT 282
DB 222 NKMQIAEGDITKVAAGDISFTLISEPDT--PIKAKLDSVDGGLTTMSSGGVNSTDTAS 279
OY 283 -----LRAAVPNDQNTILMPGLYRVLMDOVAVDNAFVVPQAAV--TRGAKDTVMITVNAOG 335
DB 280 NAVVYVARSFVNPDPGKATGMTQNTQVEIDGVKNVLLIPSLTVKNRGKAFVRLGADG 339
OY 336 GMEPREVTVAAQOGTNNIVTSGLKDGDKVNVVEGISIA 372
DB 340 KAVEREIRTGKMSMNTKESGLKEGDKVVISETITAA 376

```

RESULT 14

AA74869 ID AA74869 standard; Protein; 392 AA.

XX AC AA74869;

XX DT 21-MAR-2000 (first entry)

XX DE *Neisseria gonorrhoeae* ORF 290 protein sequence SEQ ID NO:1212.

XX KW *Neisseria meningitidis*; *Neisseria gonorrhoeae*; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
 XX antibacterial; gene therapy.

XX OS *Neisseria gonorrhoeae*.

XX PN W09957280-A2.

XX PD 11-NOV-1999.

XX PF 30-APR-1999; 99WO-US09346.

XX PR 01-MAY-1998; 98US-0083758.

XX PR 31-JUL-1998; 98US-0094869.

XX PR 02-SEP-1998; 98US-0098994.

XX PR 02-SEP-1998; 98US-0099062.

XX PR 09-OCT-1998; 98US-0103749.

XX PR 09-OCT-1998; 98US-0103794.

XX PR 09-OCT-1998; 98US-0103796.

XX PR 25-FEB-1999; 99US-0121528.

XX PA (CHIR) CHIRON CORP.

XX PA (GENO-) INST GENOMIC RES.

XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX
 DR WPI: 2000-062150/05.
 DR N-ESDB; AA253631.

PT Novel *Neisseria* polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics
 XX
 PS Claim 2; page 676; 1453pp; English.

CC AA253015 to AA254536, AA254577 to AA254615, and AA74253 to AA75941
 CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA254573 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC *Neisseria* bacteria (e.g. meningitis and septicaemia), to detect the
 CC presence of *Neisseria* bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.

XX SQ Sequence 392 AA;

Query Match 15.3%; Score 308; DB 21; Length 392;

Best Local Similarity 27.7%; Pred. No. 1.3e-18;
 Matches 110; Conservative 63; Mismatches 170; Indels 54; Gaps 10;

```

OY 7 KAMRAAALAAVALVLSGCKGDAAGGAPAGREAPAPVGVTVHPQVALTYELPCR 66
DB 3 KMKKAAVAALVAAA-----AWGGMWYLYKPEPOAAVITTAVRGDISRTVSATGE 52
OY 67 LESLRTADVRAQVGIIQKRLFOEGSYVRAGOPLYOIDSSTY-----EANLESARAOL 119
DB 53 ISPSNLVSVGAQASGQIKLVKLGQGVKKGDLAEINSTTQNTIDMEKSKLETYQAKL 112
OY 120 ATAQATLAKADADLARYKPLVAEAVSRQEVDAAVTKRSAGAKVAAQAIAKSGINLN 179
DB 113 VSAQIALGSAEKKYKROAALMKDDATSKEDLESADALAAKAVAAELKALIRQKISIN 172
OY 180 R-----SRTAPISGFIGOSKVSSEGLTNAAGDTVLATITROTNPVYVNTQSASEVWK 232
DB 173 TAESDLGYTRITATMDGTVVAIPVEEGQTVNAA-----QSTPTTVQLAN--LDMWL 221
OY 233 LRQIAEGKL--LAADGVAVGIRKFPDGTVPYPERGRLLFADPVV-----NESTGOIT 282
DB 222 NKMQIAEGDITKVAAGDISFTLISEPDT--PIKAKLDSVDGGLTTMSSGGVNSTDTAS 279
OY 283 -----LRAAVPNDQNTILMPGLYRVLMDOVAVDNAFVVPQAAV--TRGAKDTVMITVNAOG 335
DB 280 NAVVYVARSFVNPDPGKATGMTQNTQVEIDGVKNVLLIPSLTVKNRGKAFVRLGADG 339
OY 336 GMEPREVTVAAQOGTNNIVTSGLKDGDKVNVVEGISIA 372
DB 340 KAVEREIRTGKMSMNTKESGLKEGDKVVISETITAA 376

```

RESULT 15

AA74871 ID AA74871 standard; Protein; 392 AA.

XX AC AA74871;

XX DT 21-MAR-2000 (first entry)

XX DE *Neisseria meningitidis* ORF 290 protein sequence SEQ ID NO:1216.

XX DE *Neisseria meningitidis*; *Neisseria gonorrhoeae*; antigen; vaccine;

XX KW *Neisseria meningitidis*; *Neisseria gonorrhoeae*; antigen; vaccine;

```

Db      173 TAESDLGYTRITATMDGTVAIIPVEEGQTNA-----QSTPTIVQLAN--LDMWL 221
Oy      233 LRQIAEGKL--LAADGVIAVGKFDGTYVPEKGRLLFADPVV-----NESTGQIT 282
Db      222 NKMOIAEGDITKVKAGODISFTLISEPDT--PIKAKLDSVDPGLTMSGGVNSTDTAS 279
Oy      283 -----LRAAVPNQNIIMPGLYVRVLMDOVAVDNAFVPOQAV--TRGAKDTVMIVNAOG 335
Db      280 NAVYYIARSPFVNPDPGLATGTTONTVEIDGVKNVLIIPSLTYKRNKGKAFVRLGADG 339
Oy      336 GMEPREVTAAOQOGTNMIVTSGLKDGDKRVVEGISIA 372
Db      340 KAVEREIRTGKMDSMNTEVKSGLKEGDKVISEITAA 376

RESULT 12
AAV38883
ID      AAV38883 standard; Protein; 392 AA.
XX
AC      AAY38883;
XX
DT      08-OCT-1999 (first entry)
DE      Neisseria meningitidis strain A antigen encoded by ORF85.
XX
KM      Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX      treatment; Neisseria infection; meningitis; septicemia; gonorrhea.
XX
OS      Neisseria meningitidis.
XX
PN      MO9924578-A2.
XX
PD      20-MAY-1999.
XX
PF      09-OCT-1998; 98WO-IB01665.
XX
PR      01-SEP-1998; 98GB-0019016.
PR      06-NOV-1997; 97GB-0023516.
PR      14-NOV-1997; 97GB-0024190.
PR      18-NOV-1997; 97GB-0024386.
PR      27-NOV-1997; 97GB-0025158.
PR      10-DEC-1997; 97GB-0026147.
PR      14-JAN-1998; 98GB-0000759.
XX
PA      (CHIR-) CHIRON SPA.
XX
PI      Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;
XX
DR      MPI; 1999-327407/27.
XX      N-PSDB; AA212305.
XX
PT      Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
XX      diagnosis, treatment and prevention of infection
XX
PS      Claim 4; Page 428-429; 524pp; English.
XX
CC      Amino acid sequences AAV38499-Y38944 represent Neisseria meningitidis
XX      and N. gonorrhoeae antigenic proteins. They are encoded by open
XX      reading frames (ORFs) AA21972-Z1358. The antigenic proteins,
XX      their fragments, their nucleic acids and antibodies are used for
XX      diagnosis, prevention (as vaccines) or treatment of Neisseria
XX      infections, such as meningitis, septicemia and gonorrhea. Both
XX      organisms are closely related. Fragments of the nucleic acids
XX      are useful as hybridisation probes and antisense reagents.
XX
SQ      Sequence 392 AA;

Query Match      15.3%; Score 308; DB 20; Length 392;
Best Local Similarity 27.5%; Pred. No. 1.3e-18;
Matches 109; Conservative 65; Mismatches 169; Indels 54; Gaps 10;

Oy      7 KAMRAALAAVALVLSGCKGDAAGGQAPGREAPAVGVTVHPQTALTVLELPGR 66

```

```

Db      3 KMMKMAVVAALAA-----AVWGMSYLPKEFOAAVITETVRGDISRTVSARGE 52
Oy      67 LESIRTDVRAQVGGITOKRLFQSGSYVRAGQPIYQDS-----STYEANLESARQL 119
Db      53 ISPSNLVSGAQASGQIKKIVYKLGQOVKKGDLAEINISQNTLNTKSKLETYQAKL 112
Oy      120 ATAQATLAKADADLARFKPVAAEAASROEYDAVTAKRSEAGVKAQAIAKSGAINLN 179
Db      113 VSAQIALGSAEKYKRRQALMKODATAKEDESNQDLAAKAVNAELKILRQSKISIN 172
Oy      180 R-----SRITADISGFQGSVSEGTLLNAGDTVLATIRQTNPMYVNTQSASEVMK 232
Db      173 TAESDLGYTRITATMDGTVAIIPVEEGQTNA-----QSTPTIVQLAN--LDMWL 221
Oy      233 LRQIAEGKL--LAADGVIAVGKFDGTYVPEKGRLLFADPVV-----NESTGQIT 282
Db      222 NKMOIAEGDITKVKAGODISFTLISEPDT--PIKAKLDSVDPGLTMSGGVNSTDTAS 279
Oy      283 -----LRAAVPNQNIIMPGLYVRVLMDOVAVDNAFVPOQAV--TRGAKDTVMIVNAOG 335
Db      280 NAVYYIARSPFVNPDPGLATGTTONTVEIDGVKNVLIIPSLTYKRNKGKAFVRLGADG 339
Oy      336 GMEPREVTAAOQOGTNMIVTSGLKDGDKRVVEGISIA 372
Db      340 KAVEREIRTGKMDSMNTEVKSGLKEGDKVISEITAA 376

RESULT 13
AAV38884
ID      AAV38884 standard; Protein; 392 AA.
XX
AC      AAY38884;
XX
DT      08-OCT-1999 (first entry)
DE      Neisseria gonorrhoeae antigenic protein encoded by ORF85.
XX
KM      Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX      treatment; Neisseria infection; meningitis; septicemia; gonorrhea.
XX
OS      Neisseria gonorrhoeae.
XX
PN      MO9924578-A2.
XX
PD      20-MAY-1999.
XX
PF      09-OCT-1998; 98WO-IB01665.
XX
PR      01-SEP-1998; 98GB-0019016.
PR      06-NOV-1997; 97GB-0023516.
PR      14-NOV-1997; 97GB-0024190.
PR      18-NOV-1997; 97GB-0024386.
PR      27-NOV-1997; 97GB-0025158.
PR      10-DEC-1997; 97GB-0026147.
PR      14-JAN-1998; 98GB-0000759.
XX
PA      (CHIR-) CHIRON SPA.
XX
PI      Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;
XX
DR      MPI; 1999-327407/27.
XX      N-PSDB; AA212306.
XX
PT      Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
XX      diagnosis, treatment and prevention of infection
XX
PS      Claim 4; Page 430; 524pp; English.
XX
CC      Amino acid sequences AAV38499-Y38944 represent Neisseria meningitidis
XX      and N. gonorrhoeae antigenic proteins. They are encoded by open
XX      reading frames (ORFs) AA21972-Z1358. The antigenic proteins,
XX      their fragments, their nucleic acids and antibodies are used for

```


PN MO3927129-A1.
XX
PD 03-JUN-1999.
XX
PF 25-NOV-1998; 98WO-US25247.
XX
PR 25-NOV-1997; 97US-0066517.
XX
PA (GENO) GEN HOSPITAL CORP.
XX
PI Ausubel F, Cao H, Drenkard E, Goodman HM, Mahajan-Miklos S;
PI Rahme LG, Tan M, Tsongalis J;
XX
XX WPI, 1999-357851/30.
DR
XX Virulence factors useful in developing disease treatments
PT
XX
PS Disclosure; Fig 14B; 228pp; English.
XX
XX The present sequence represents a *Pseudomonas aeruginosa* polypeptide
CC sequence. *P. aeruginosa* is an opportunistic human pathogen present in
CC soil water and plants. The specification describes virulence polypeptides
CC and nucleic acid sequence encoding such polypeptides. These sequences
CC can be used to identify a compound which is capable of decreasing the
CC expression of a pathogenic virulence factor. Compounds that inhibit
CC the expression or activity of virulence factor polypeptides can be
CC used to treat pathogenic infections, especially where the infection
CC is a *P. aeruginosa* infection.
CC note: the sequences given in the specification were poorly legible, and
CC in some instances assumptions were made as to the identity of the
CC residue; it is therefore possible that the sequence given below is
CC not entirely correct.
XX
SQ Sequence 202 AA;

Query Match 18.4%; Score 370.5; DB 20; Length 202;
Best Local Similarity 37.9%; Pred. No. 1.5e-24;
Matches 80; Conservative 43; Mismatches 73; Indels 15; Gaps 5;

QY 205 GDTVLATIRQTNPMYVNTQASAEVWKLKROIAEGKL-LAADGVIANGIRFDDGTVPE 263
DB 1 GQANAMATVQQLDPIYVDVTPSTALMRRELASGOLERAGDNAAKSLKLEDSQYPL 60
QY 264 KGRLLFADPVNVESTGOITIRAAVNDONILMPGLVYRVLMDOQVAVDNAFVVPQAVTRG 323
DB 61 EGRLEFSEVSDGEGSVTIRAVFPNNNELPGMFVAHQLEGGVKQRAIILAPQCGVTRD 120
QY 324 AKD--TNWIVNAQGMPEPEVTVAAQOGTNWITSGLDKDGKVVVEGISIA--GITGAKK 379
DB 121 LKGGATLAVNNAQNQVLEKRVIKADVIGDKMLVTGSLNAGDKILIREGIQFQPGVE-VKT 179
QY 380 VTPKEWASSENOAAPPQSGVQTSAPKTPASE 410
DB 180 VPARKVVASAQKADAP-----AKTDSK 201

RESULT 9

ABGI7698
ID ABGI7698 standard; Protein; 703 AA.

AC ABGI7698;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #17689.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

XX WO200175067-A2.

PD 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
PF
XX 31-MAR-2000; 2000US-0540217.
PR
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI, 2001-639362/73.
DR
XX N-PSDB; AAS81885.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensic, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 20; SEQ ID No 48057; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridization probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensic, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 703 AA;

Query Match 16.8%; Score 339; DB 22; Length 703;
Best Local Similarity 28.7%; Pred. No. 5.6e-21;
Matches 87; Conservative 62; Mismatches 110; Indels 44; Gaps 5;

QY 84 QKRLFOEGSYRAGOPFYQISSTYEANLESARAQLATAQATLAKADA-----DLARY 136
DB 386 EEGISQIRAYDSRQQLAVNDTQGHETREYENAAATYDSAVICAPAAANIAQLTVRY 445
QY 137 KPIVAEAESRQEDAAVTAKRSAGVKAQAALIKSAGINLNSRTAPISGFIGQSKV 196
DB 446 QKLGTYISKQFVQDLADQANAAVTAKAAVETKQINLTPK--APRLVAALSNV 503
QY 197 SEGTLNAGDTVLTATIRQTNPMYVNTQASAEVWKLKROIAEGKLAAADVIANGIRFD 256
DB 504 TEGALVQNGQATATATVQQLDPIYVDVQSSNDPRLKQBLANGTLKQENCAKAVSLITS 563
QY 257 DGTVPKGRLLFADPVNVESTGOITLRAAVNDONILMPGLVYRVLMDOQVAVDNAFVVP 316
DB 564 DGIRFPQGTLEFSDVYVQRTGSITLRAIRPNPDHLLP----- 604
QY 317 QQAVTRGAKDTVMVNAQGMPE--REVTVAAQOGTNWITSGIKD--GDKVVVEGIS 370
DB 605 -----NIDEMNKLAGEFETNHRIRPLISIGTRWVLQTMWQGNATAGRRLIPSS 654
QY 371 IAG 373
DB 655 VSG 657

CC identified as being required for bacterial growth and proliferation, can
CC be used for antisense therapy for killing bacteria.

CC Sequence 464 AA;

Query Match 21.4%; Score 432; DB 21; Length 464;

Best Local Similarity 32.2%; Pred. No. 1.9e-29;

Matches 119; Conservative 63; Mismatches 168; Indels 20; Gaps 6;

CC 34 GGAPGAREAPVAVGVVTHPQTVALVELPGRLESRTADYRAQVGGIIQKRLFOEGSY 93
DB 98 GGRGMRSGPLAPVQATAVEQAVPRYLGLGTTAANTVTRSRVDGLIALHFOEGQ 157
QY 94 VRAQPLVQIDSSSTEANIESAPQATLAKADADLARYKPLVAEAVSROEYDA 153
DB 158 VVAGGLAEIDPSQFKVALAQOGLAKKATLANARRLARYOOLAKTNLVSROELDAQ 217
QY 154 VTAKSAEAGVKAQAATKSAGININRSRTAPISGFIGSVSEGTLLNADPTVLAT 213
DB 218 QALVSETEGTTKADASVASAQLQDWSRTAPVGRVGLKQVDVGNQISSGDTTGIYVI 277
QY 214 RQTNPMYVNTQSAEVMKLRQIAEGKLADGVIAVGI-KPDDGTVYPERGRLLFPDP 272
DB 278 TQTHPIDVFTLPESDIATVVOAKAGKPLVVEAMDRTNSKLSGTF-----LLSIDN 330
QY 273 VNNESTGOITLRAAVPNDQNTLMPGLYV--RVLMDOVAVDNAFVVPQAVTRGAK-DTVM 329
DB 331 QIDATGTITKVARFNNQDDALFPNQFVNARMLVD--TEQNAVVIPTALQMGNGHFVW 388
QY 330 IVNAQGMPEPREVTVAAQOQGTNMTVTSGLKDGKRVVVEGI-----SIAGITGAKKVT 382
DB 389 VLNSNKYSKHLVTPGIDPSQKVIIRAGISAGDRVVTGIDRLTEGAKVEVEAQSATTP 448
QY 383 KEMASSENQA 392
DB 449 EKAATREYA 458

RESULT 7
AAG98971
ID AAG98971 standard; Protein; 464 AA.
AC AAG98971;
XX
DT 26-SEP-2001 (first entry)
DE E. coli growth and proliferation related protein sequence SEQ ID NO:441.
XX
KW Escherichia coli; growth; proliferation; microbial; antimicrobial;
KW bacterial infection; microorganism.
XX
OS Escherichia coli.
XX
EN WO200134810-A2.
XX
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000WO-US30950.
XX
PR 09-NOV-1999; 99US-0164415.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Forsyth RA, Ohlsen K, Zykkind J;
XX
DR WPI; 2001-335933/35.
XX
DR N-PSDB; AAH84642.
XX
PT Novel nucleic acids that inhibit Escherichia coli proliferation, useful
PT for screening for homologous genes and for designing expression vectors
XX
PS Claim 19; Page 496-497; 522pp; English.

XX AAH84373 to AAH84499 represent Escherichia coli growth and proliferation
CC related DNA sequences (1). AAH84500 to AAH84670 encode the E. coli
CC growth and proliferation related proteins given in AAG99078 and AAG98830
CC to AAG98999. (1) can be used as potential targets for the generation of
CC new antimicrobial agents, and for identification of compounds which
CC interact with the gene products of (1). In addition the expression of
CC (1) and the purification of the proteins, the purified proteins can be
CC used to generate reagents and screen small molecule libraries or other
CC candidate compound libraries for compounds that can be further developed
CC to yield novel antimicrobial compounds. In addition, nucleic acid probes
CC complementary to (1) that are specific for particular species of
CC microorganisms can be used to identify particular microorganism species
CC in clinical specimens, therefore, providing a rapid and dependable
CC method by which to identify the causative agents of a bacterial
CC infection. Also, antibodies generated against proteins translated from
CC mRNA transcribed from proliferation-related sequences can also be used
CC to screen for specific microorganisms that produce such proteins in a
CC species-specific manner. AAH84371 and AAH84670 represent sequencing
CC primers used in the isolation of E. coli growth and proliferation
CC related sequence, which are used in an example from the present
CC invention.

CC Sequence 464 AA;

Query Match 21.4%; Score 432; DB 22; Length 464;

Best Local Similarity 32.2%; Pred. No. 1.9e-29;

Matches 119; Conservative 63; Mismatches 168; Indels 20; Gaps 6;

QY 34 GGAPGAREAPVAVGVVTHPQTVALVELPGRLESRTADYRAQVGGIIQKRLFOEGSY 93
DB 98 GGRGMRSGPLAPVQATAVEQAVPRYLGLGTTAANTVTRSRVDGLIALHFOEGQ 157
QY 94 VRAQPLVQIDSSSTEANIESAPQATLAKADADLARYKPLVAEAVSROEYDA 153
DB 158 VVAGGLAEIDPSQFKVALAQOGLAKKATLANARRLARYOOLAKTNLVSROELDAQ 217
QY 154 VTAKSAEAGVKAQAATKSAGININRSRTAPISGFIGSVSEGTLLNADPTVLAT 213
DB 218 QALVSETEGTTKADASVASAQLQDWSRTAPVGRVGLKQVDVGNQISSGDTTGIYVI 277
QY 214 RQTNPMYVNTQSAEVMKLRQIAEGKLADGVIAVGI-KPDDGTVYPERGRLLFPDP 272
DB 278 TQTHPIDVFTLPESDIATVVOAKAGKPLVVEAMDRTNSKLSGTF-----LLSIDN 330
QY 273 VNNESTGOITLRAAVPNDQNTLMPGLYV--RVLMDOVAVDNAFVVPQAVTRGAK-DTVM 329
DB 331 QIDATGTITKVARFNNQDDALFPNQFVNARMLVD--TEQNAVVIPTALQMGNGHFVW 388
QY 330 IVNAQGMPEPREVTVAAQOQGTNMTVTSGLKDGKRVVVEGI-----SIAGITGAKKVT 382
DB 389 VLNSNKYSKHLVTPGIDPSQKVIIRAGISAGDRVVTGIDRLTEGAKVEVEAQSATTP 448
QY 383 KEMASSENQA 392
DB 449 EKAATREYA 458

RESULT 8
AAV29280
ID AAV29280 standard; Protein; 202 AA.
AC AAV29280;
XX
DT 25-OCT-1999 (first entry)
DE Partial amino acid sequence of PA14 mexA.
XX
KW Human pathogen; virulence polypeptide; virulence factor;
KW pathogenic infection; Pseudomonas aeruginosa infection.
XX
OS Pseudomonas aeruginosa.

ABG15606
ID ABG15606 standard; Protein; 978 AA.
XX
XX AC
XX ABG15606;
XX
XX
XX DT 18-FEB-2002 (first entry)
XX
XX DE Novel human diagnostic protein #15597.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX OS Homo sapiens.
XX
XX MN WO200175067-A2.
XX
XX PD 11-OCT-2001.
XX
XX PF 30-MAR-2001; 2001WO-US08631.
XX
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX
XX PA (HYSE-) HYSEQ INC.
XX
XX PI Drmanac RT, Liu C, Tang YF;
XX
XX WPI, 2001-639362/73.
XX DR N-PSDB; AAS797933.
XX
XX
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity -
XX
XX
XX PS Claim 20; SEQ ID No 45965; 103pp; English.
XX
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridization probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human
XX CC diagnostic amino acid sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pat_sequences.
XX
XX
XX Sequence 978 AA;
XX

```

QY      105 SSTYEANIESBARQAUTATQATILAK--ABDDILA-----RKPLVAAAVASRQEX 150
Db      406 TRAYPCPLTPPIPAAGHPEHAHCDFDELSLMPFNKDIHLFRRAKYVA-----FW 459
QY      151 DAAVTAKRSAAEGYKAA-----QAAIKSAGINIMNSRTAPISGFJGSGKVSEGTLLMAG 2050
Db      460 RIASSSSSHASKRRLRAISPCTQPVPVTRIRIYLATKTSPISGRIGKSNVTEGALVONG 5119
QY      206 DTTVALITRQTNPBMVNANTOSASEVMKLRRQIAEGLTLADGVIAVGIFEDDVTYPBKG 2655
Db      520 CATALATATQOGLPIVDYDTQGSNDPLRLKQELANGTLKXENGGAKKSILTSDGIKRPDG 5799
QY      266 RILEPADPVNVESTGOITLRPAVPNDONITIMPGLYYRVLMDOYAVDNAFVVPQOAVTREGAK 3255
Db      580 TLEFSDVAVDQOTGISTILLRAIPPNPDHTLLPDMFYARLEEGLNPMNAIIIVPOOGVTRTPR 6399
QY      366 D--TYMTIYNAGCGMFPREVTYAQQOOGSTWTVTSGLKDGNKKVVVEGISIA-----GITGA 3777
Db      640 GDATAVLVVGADDKVETREPIVASQAIGDKMLVTEGLKAGDRVVISGLQKVRPGVQPGIMMA 6999
QY      378 KKVT--PKEMA 386
Db      700 EKULSELPSAMS 710

```

RESULT 6
 AAB15924
 ID AAB15924 standard; Protein; 464 AA.
 XX
 AC AAB15924;
 XX
 DT 05-OCT-2000 (first entry)
 XX
 DE E. coli proliferation associated protein sequence SEQ ID NO:281.
 XX
 KW Escherichia coli; E. coli; proliferation; inhibition; screening;
 XX antimicrobial; bacterial growth; antisense therapy; antibacterial.
 XX
 OS Escherichia coli.
 XX
 PN WO200044906-A2.
 XX
 PD 03-AUG-2000.
 XX
 PF 27-JAN-2000; 2000MO-US02200.
 XX
 PR 27-JAN-1999; 99US-0117405.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Zyskind J, Ohlsen KL, Trawick J, Forsyth RA, Froelich JM, Carr GJ;
 PI Yamamoto RT, Xu HH;
 XX
 DR WPI, 2000-514822/46.
 XX
 DR N-PEDB; AAA65929.
 XX
 PT Novel polynucleotides and polypeptides associated with microorganism
 PT proliferation, used to identify inhibitors of bacterial growth and
 PT proliferation, for use in antisense therapy -
 XX
 PS Claim 11; Page 205-206; 316pp; English.
 XX
 CC AAA65809 to AAA65889 and AAA66058 to AAA66138 represent nucleotide
 CC sequences derived from Escherichia coli which inhibit E. coli
 CC proliferation. AAA65890 to AAA66055 and AAB15886 to AAB16040 represent
 CC nucleotide and protein sequences associated with E. coli proliferation.
 CC AAA66056 and AAA66057 represent primers used for sequencing E. coli
 CC proliferation inhibiting nucleotide inserts in an example from the
 CC present invention. Methods from the present invention can be used to
 CC identify a proliferation- required gene in a microorganism, by contacting
 CC a microorganism with a proliferation-requried gene activity inhibitory
 CC nucleic acid identified in another organism, and determining if
 CC inhibition occurs in the second microorganism. The nucleic acid sequences

PD 10-OCT-2002.
 XX 12-FEB-2002; 2002MO-IB02069.
 XX 12-FEB-2001; 2001GB-0003424.
 XX (CHIR-) CHIRON SPA.
 XX Fontana MR, Pizza M, Masignani V, Monaci E;
 XX MPI; 2003-058415/05.
 DR N-PSDB; AB039062.
 XX
 PT New protein from *Neisseria gonorrhoeae*, useful for the manufacture of a
 PS medicament for treating or preventing *N. gonorrhoeae* infection -
 XX Disclosure; Page 386; 815pp; English.
 CC The present invention relates to proteins from *Neisseria gonorrhoeae*.
 CC Also disclosed are the nucleic acid molecules encoding the proteins and
 CC antibodies that specifically bind to the proteins. The composition
 CC comprising the protein, nucleic acid or antibody is useful for the
 CC manufacture of a medicament for treating or preventing *N. gonorrhoeae*
 CC infection, this may be in the form of a vaccine or gene therapy.
 CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
 CC molecules of the invention.
 XX
 SQ Sequence 412 AA;
 Query Match 96.2%; Score 1943; DB 24; Length 412;
 Best Local Similarity 95.6%; Pred. No. 4.9e-163;
 Matches 394; Conservative 8; Mismatches 10; Indels 0; Gaps 0;
 QY 1 MAFYAFKMRRAAALAAVALVLSCKGKGDAAQGGOPAGREAPAPVGVVTHPQTVALT 60
 DB 1 MAFYAFKMRRAAALAAVALVLSCKGKGDAAQGGOPAGREAPAPVGVVTHPQTVALT 60
 QY 61 VELPGRLESLRTADVRAQVGGIIQKRLFOEGSYVRAGOPLYQIDSTYEANLESARAQLA 120
 DB 61 VELPGRLESLRTADVRAQVGGIIQKRLFOEGSYVRAGOPLYQIDSTYEANLESARAQLA 120
 QY 121 TQAQTLAQAADADLAAYKPLVAAEAVSRQRYDAANTAKSAEAGVYAAQAALIKSAGINLR 180
 DB 121 TQAQTLAQAADADLAAYKPLVAAEAVSRQRYDAANTAKSAEAGVYAAQAALIKSAGINLR 180
 QY 121 TQAQTLAQAADADLAAYKPLVAAEAVSRQRYDAANTAKSAEAGVYAAQAALIKSAGINLR 180
 DB 121 TQAQTLAQAADADLAAYKPLVAAEAVSRQRYDAANTAKSAEAGVYAAQAALIKSAGINLR 180
 QY 181 SRITAPISGFIGOSKVSSEGTLLNAGDTTVLATIRQTNPMYVNVQSAEVMKLRQIAEG 240
 DB 181 SRITAPISGFIGOSKVSSEGTLLNAGDTTVLATIRQTNPMYVNVQSAEVMKLRQIAEG 240
 QY 241 KLLAADGVIAVGIRKDDGVYPEKGRLLFADPVVNESTGOITLRAAVENDONILMPGLYV 300
 DB 241 KLLAADGVIAVGIRKDDGVYPEKGRLLFADPVVNESTGOITLRAAVENDONILMPGLYV 300
 QY 241 KLLAADGVIAVGIRKDDGVYPEKGRLLFADPVVNESTGOITLRAAVENDONILMPGLYV 300
 DB 241 KLLAADGVIAVGIRKDDGVYPEKGRLLFADPVVNESTGOITLRAAVENDONILMPGLYV 300
 QY 301 RVLMDQVAVDVAFFVPOQAVTRGAKDTMTIYVNAQGMPEPVEVTVAAQOGTWTWITSGIKD 360
 DB 301 RVLMDQVAVDVAFFVPOQAVTRGAKDTMTIYVNAQGMPEPVEVTVAAQOGTWTWITSGIKD 360
 QY 301 RVLMDQVAVDVAFFVPOQAVTRGAKDTMTIYVNAQGMPEPVEVTVAAQOGTWTWITSGIKD 360
 DB 301 RVLMDQVAVDVAFFVPOQAVTRGAKDTMTIYVNAQGMPEPVEVTVAAQOGTWTWITSGIKD 360
 QY 361 GDKVVEGISTAGITGAKKVTPEKMASSENOAAAPQSGVOTASEKTKTASEAE 412
 DB 361 GDKVVEGISTAGITGAKKVTPEKMASSENOAAAPQSGVOTASEKTKTASEAE 412
 Db 361 GDKVVEGISTAGITGAKKVTPEKMASSENOAAAPQSGVOTASEKTKTASEAE 412
 RESULT 4
 ID AAY32854 standard; Protein; 397 AA.
 AC AAY32854;
 XX AAY32854;
 XX 01-NOV-1999 (first entry)
 DT E. coli acrA protein sequence.
 XX
 XX AcrA; acb; tolC; organic solvent resistance; indigo production;

KM steroid conversion; flooded crude oil treatment.
 XX Escherichia coli.
 OS
 XX Key Location/Qualifiers
 FH Peptide 1..24
 FT Protein /note= "signal peptide"
 FT Protein 25..397
 FT Protein /note= "mature acrA"
 XX
 XX JP11221080-A.
 XX
 XX 17-AUG-1999.
 XX
 XX 09-FEB-1998; 98JP-0027537.
 XX
 XX 09-FEB-1998; 98JP-0027537.
 XX
 XX (MEIJ) MEIJI SEIKA KAISHA LTD.
 XX
 XX MPI; 1999-520718/44.
 DR N-PSDB; AA211064.
 XX
 PT Escherichia coli having organic solvent resistance - useful for
 PT production of indigo, conversion of steroid and treating flooded
 PT crude oil
 XX
 XX
 PS Claim 5; Page 8-9; 14pp; Japanese.
 XX
 CC This sequence represents the acrA protein. The invention relates
 CC to a method for the preparation of E. coli having organic solvent
 CC resistance or of increased organic solvent resistance including a step of
 CC transforming the E. coli with at least one gene of acrA, acb and tolC.
 CC The E. coli can be used for the production of indigo from indol,
 CC conversion of a steroid such as cholesterol and treatment of flooded
 CC crude oil.
 XX
 SQ Sequence 397 AA;
 Query Match 37.8%; Score 762.5; DB 20; Length 397;
 Best Local Similarity 42.2%; Pred. No. 9.3e-59;
 Matches 168; Conservative 76; Mismatches 129; Indels 25; Gaps 6;
 QY 11 AALAAVAALVLSCKGKGDAAQGGOPAGREAPAPVGVVTHPQTVALFVELPGRLESL 70
 DB 11 AALAAVAALVLSCKGKGDAAQGGOPAGREAPAPVGVVTHPQTVALFVELPGRLESL 70
 QY 71 RTADVRAQVGGIIQKRLFOEGSYVRAGOPLYQIDSTYEANLESARAQLAQAATLAKAD 130
 DB 71 RTADVRAQVGGIIQKRLFOEGSYVRAGOPLYQIDSTYEANLESARAQLAQAATLAKAD 130
 QY 64 RIAEVRQVSGGIIKRNFKESGSDIEAGVSLYQIDPATYQATYDSAKGDLAKAQAANVAIQ 123
 DB 64 RIAEVRQVSGGIIKRNFKESGSDIEAGVSLYQIDPATYQATYDSAKGDLAKAQAANVAIQ 123
 QY 131 ADLARYEPVAAEAVSRQRYDAANTAKSAEAGVYAAQAALIKSAGINLR 190
 DB 131 ADLARYEPVAAEAVSRQRYDAANTAKSAEAGVYAAQAALIKSAGINLR 190
 QY 124 LTVNRYPKLLGTQYISKQETDQALADAAQNAAYTAKAAVEITRINLAATKVTSPISGR 183
 DB 124 LTVNRYPKLLGTQYISKQETDQALADAAQNAAYTAKAAVEITRINLAATKVTSPISGR 183
 QY 191 IGOSKVSSEGTLLNAGDTTVLATIRQTNPMYVNVQSAEVMKLRQIAEGKLLAADGVIA 250
 DB 191 IGOSKVSSEGTLLNAGDTTVLATIRQTNPMYVNVQSAEVMKLRQIAEGKLLAADGVIA 250
 QY 184 IGKSNVTEGALVQNGQATLALATVQQLDPIYDVDTQSSNDPLRLKQELANGTLKENGAK 243
 DB 184 IGKSNVTEGALVQNGQATLALATVQQLDPIYDVDTQSSNDPLRLKQELANGTLKENGAK 243
 QY 251 VGKFDGDTVYPEKGRLLFADPVVNESTGOITLRAAVENDONILMPGLYVRLMDQVAVD 310
 DB 251 VGKFDGDTVYPEKGRLLFADPVVNESTGOITLRAAVENDONILMPGLYVRLMDQVAVD 310
 QY 244 VSLITSDGIRKFPQDGTLEFSDVTVDTGTSITLRAIPFPNDHTLLPMPFARLREGLNP 303
 DB 244 VSLITSDGIRKFPQDGTLEFSDVTVDTGTSITLRAIPFPNDHTLLPMPFARLREGLNP 303
 QY 311 NAFVPOQAVTRGAKD--TMTIYVNAQGMPEPVEVTVAAQOGTWTWITSGIKDGRVVEG 368
 DB 311 NAFVPOQAVTRGAKD--TMTIYVNAQGMPEPVEVTVAAQOGTWTWITSGIKDGRVVEG 368
 QY 304 NALIVPOQGVTRTPRGDVLVVGADDKVETRPVIAQIGDKLVTEGKAGGRVVISG 363
 DB 304 NALIVPOQGVTRTPRGDVLVVGADDKVETRPVIAQIGDKLVTEGKAGGRVVISG 363
 QY 369 ISLIGITGAKKVTPEKMASSENOAAAPQSGV 400
 DB 364 L-----QKVRPGVQVKAQEVTAIDNNQQA--SGAQ 391
 RESULT 5

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2003, 13:52:56 ; Search time 44 Seconds

(without alignments)
1486.258 Million cell updates/sec

Title: US-09-889-756A-2

Perfect score: 2019
Sequence: 1 MAFYAFKAMRAAALAAVAL.....AAPGSGVOTASEAKTASAE 412

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_19Jun03:*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2019	100.0	412	21	AAV90283
2	1989	98.5	412	23	AAU91064
3	1943	96.2	412	24	ABP78092
4	762.5	37.8	397	20	AAV32854
5	459.5	22.8	978	22	ABG15606
6	432	21.4	464	21	AAAB15924
7	432	21.4	464	22	AAAG96971
8	370.5	18.4	202	20	AAV29280
9	339	16.8	703	22	ABG17698

10	320.5	15.9	627	22	ABG18262
11	313	15.5	392	24	ABP79725
12	308	15.3	392	20	AAV38883
13	308	15.3	392	20	AAV38884
14	308	15.3	392	21	AAV74869
15	308	15.3	392	21	AAV74871
16	308	15.3	392	23	AAU72963
17	308	15.3	392	23	AAU73013
18	271	13.4	334	20	AAV38882
19	271	13.4	334	21	AAV74870
20	217	10.7	102	23	ABP31381
21	168	9.3	394	24	ABP79084
22	175	8.7	659	22	ABP3061
23	169	8.4	309	22	AAU34575
24	169	8.4	309	22	AAU38257
25	165.5	8.2	89	23	ABP04272
26	164.5	8.1	175	21	AAAB44578
27	164.5	8.1	175	23	ABP54530
28	161.5	8.0	285	22	AAU36320
29	158	7.8	115	23	ABP31959
30	158	7.8	262	23	AAAG80026
31	156.5	7.8	390	22	AAU35553
32	156.5	7.8	399	24	AAE30466
33	151.5	7.5	399	21	AAE81743
34	151.5	7.5	399	24	ABU01173
35	151	7.5	645	22	ABG21616
36	140	6.9	684	22	ABG17838
37	138.5	6.9	355	22	AAAG98387
38	137.5	6.8	444	20	AAV55920
39	133	6.6	1201	22	ABB61629
40	132.5	6.6	392	20	AAV38881
41	132	6.6	387	23	AAO17576
42	130	6.4	2453	21	AAAB12454
43	129.5	6.4	512	21	AAAG31336
44	129	6.4	4572	19	AAW52845
45	127.5	6.3	329	19	AAW71499

ALIGNMENTS

RESULT 1
AAV90283 standard; Protein: 412 AA.

AC AAV90283;
24-OCT-2000 (first entry)

DE N. meningitidis BASH055 protein sequence.

KW BASH055; diagnosis; microbial infection; invasive bacterial disease;
KW Neisseria meningitidis infection; upper respiratory tract infection;
KW bacteraemia; meningitis; therapy.

OS Neisseria meningitidis.
PN WO2000043517-A1.

PD 27-JUL-2000.
PF 19-JAN-2000; 2000WO-EP00425.
PR 22-JAN-1999; 99GB-0001462.
PR 29-JAN-1999; 99GB-0002069.

(SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
Thomnard J;
WPI: 2000-476199/41.
DR N-PSDB; AAA37645.

Novel human diagno
N. gonorrhoeae ami
Neisseria meningit
Neisseria gonorrhoe
Neisseria meningit
Neisseria meningit
Neisseria meningit
Neisseria meningit
Neisseria meningit
Human glycoprotein
N. gonorrhoeae ami
C. glutamicum prote
E. coli cellular p
Salmonella typhi c
Human ORF32 prote
Virulence gene pro
Actinobacillus ple
Pseudomonas aerugi
Human ORF32 prote
A. cryptum Alub pr
Haemophilus influe
H. influenzae emrA
Streptococcus pneu
S. pneumoniae type
Novel human diagno
Novel human diagno
Escherichia coli p
Pseudomonas fluore
Drosophila melanog
Neisseria meningit
M. catarrhalis MCAL
HNRCR protein sequ
Arabidopsis thaliana
A. mediterranei ri
Helicobacter polyp

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR	EMBL: X114199; CAAB32413.1; -.
DR	PIR: S02387; BVBRCD.
DR	InterPro: IPR006143; HLYD.
DR	InterPro: IPR006144; HLYD_FAMILY.
DR	InterPro: IPR003997; RTX.D.
DR	PFAM: PF00529; HLYD.1.
DR	PRINTS: PR01490; RTXTOXIND.
DR	PROSITE: PS00643; HLYD_FAMILY.1.
DR	Hemolysis; Transpot; Transmembrane; Inner membrane.
KW	DOMAIN 1 55
FT	DOMAIN 56 75
FT	TRANSMEM 56 75
FT	DOMAIN 76 440
SC	SEQUENCE 440 AA; 48011 MW; B786328A92DD048 CRC64.

AC Q52969; 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical sensor-like histidine kinase R01002 (EC 2.7.3.-).
 GN R01002 OR SMC00059.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
 NC NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RC MEDLINE=21396507; PubMed=11481430;
 RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
 RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
 RA Gordin T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masny D.,
 RA Pohl T., Portetelle D., Puehler A., Fumelle B., Rampeger U.,
 RA Renard C., Thebaud P., Vandenbol M., Weidner S., Gallibert F.,
 RT "Analysis of the chromosome sequence of the legume symbiont
 RT Sinorhizobium meliloti strain 1021."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
 RN [2]
 RP SEQUENCE OF 1-87 FROM N.A.
 RC STRAIN=RCR2011 / S047;
 RC MEDLINE=95276304; PubMed=7756693;
 RA Keller M., Roxlau A., Weng W.M., Schmidt M., Quandt J., Niehaus K.,
 RA Jording D., Arnold W., Puehler A.;
 RT "Molecular analysis of the Rhizobium meliloti mucr gene regulating the
 RT biosynthesis of the exopolysaccharides succinoglycan and
 RT galactoglycan."
 RL Mol. Plant Microbe Interact. 8:267-277(1995).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: Contains 1 histidine kinase domain.
 CC -1- SIMILARITY: Contains 1 PAS-associated C-terminal (PAC) domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AL591785; CA045574.1; -;
 DR EMBL; L37353; AAA74242.1; -;
 DR InterPro; IPR003594; ATPbind_ATPase.
 DR InterPro; IPR004358; Bact_sens_pr_C.
 DR InterPro; IPR003661; His_kinase.
 DR InterPro; IPR005467; His_kinase.
 DR InterPro; IPR001610; PAC.
 DR InterPro; IPR000700; PAS-assoc_C.
 DR Pfam; PF02518; HATPase_C; 1.
 DR Pfam; PF00512; HisKA; 1.
 DR PRINTS; PR00344; BCTR1SENSOR.
 DR SMART; SM00387; HATPase_C; 1.
 DR SMART; SM00388; HisKA; 1.
 DR SMART; SM00086; HisKA; 1.
 DR PROSITE; PS50109; HIS_KIN; 1.
 DR PROSITE; PS50113; PAC; 1.
 KW Hypothetical protein; Sensory transduction; Transferase; Kinase;
 KW Transmembrane; Phosphorylation; Complete proteome.
 FT TRANSMEM 36
 FT TRANSMEM 56
 FT TRANSMEM 61
 FT DOMAIN 173 228
 FT DOMAIN 246 466
 FT MOD RES 249 249
 SO SEQUENCE 525 AA; 55899 MW; 5366357A63F99F0 CRC64;
 Query Match 6.1%; Score 123; DB 1; Length 525;
 Best Local Similarity 22.1%; Pred. No. 2.2;
 Matches 120; Conservative 72; Mismatches 174; Indels 176; Gaps 27;

QY 1 MAFYAFKAMRAA-----IAAAVALVSSCGKGD-----AAGCGQPAQR 40
 DB 26 IARDFFRLRAVAAVSLTALPIVLFALTLALPVSAALPAGAAALWASASLLAAAAAIAAGR 85
 QY 41 E-----APAP-----VGVTVHQTVALTYELGR-----LESRTADV 75
 DB 86 EPTDGEVEFAPAPLPLDNLNAYDLFAGLVTVH-DTRGHVLSVGRDASEYLKLMRDHG 144
 QY 76 RAQVGIIOKRLFOEGSVVRAGQPLYQIDSSSTYEANESAPQAQATATQATLAKADADLAR 135
 DB 145 R---GFIEQIHVSORIAFLRA-----IDS-----LRDSEK---SAVDIRLEKTSADCPQ 188
 QY 136 Y-----KPLVAE-----AVSRQEDPAAVTAKSAEAGVAAQAATKAGINLRSRIT 184
 DB 189 FAHYICENTPLRDAEGNLLAIVAQSRDVEEARLQAEAAKAAHA--ESA--NDAKTRFL 244
 QY 185 A-----PLSGITGQSKVSEG-----TLNAGDTYLAITRQ-----215
 DB 245 AAVSHELRTPLNALITGSDVLAGEYFGKLENDROREYVSLIHQSGTHLSVVTMLDMSK 304
 QY 216 -----TTPMTVNTQSASEVW-----KLRQIAEGKLAAD-----246
 DB 305 IEAGRYELLEPRVVAEAIACEMLSHOAEKGVRLTSRVSVC---INADQRAQ 361
 QY 247 ---GVIAVGIKFDGTYVPEKGRLLFADPVVNESTGQITLRAAVPNDQNIIMPGLYVRVL 303
 DB 362 ILINLIGNAIKFTD-----RGLVTVDAALLEGMLKLTV-----SDTGICIAADKIQML 410
 QY 304 MDQVAVDNAPVYPOQAVTRGAKDT---VMYVNAQSGHEPRVTV--AQOQGTNMIYVSG 358
 DB 411 -----GQPFVQIONDYTRRYEGTGLSLVGLAEHLGGDVSTRSAGECTVIVWTTP- 463
 QY 359 KDQGVVVEGISIGITGAKVTPEKWAASSENQAAP-----QSGVQTASRAK 406
 DB 464 SDASGAARQCADAPVTEPPRLKQHADGKREAGVPLSEALHTGELGREGHGAAQAK 523
 QY 407 TA 408
 DB 524 TA 525
 RESULT 40
 CYAD BORPE STANDARD; PRT; 440 AA.
 ID CYAD BORPE.
 AC P11091;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Cyad protein.
 GN CYAD.
 OS Bordetella pertussis.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; Bordetella.
 OX NCBI_TaxID=520;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=18323;
 RC MEDLINE=89091151; PubMed=2905265;
 RX Glaser P., Sakamoto H., Bellalou J., Ullmann A., Danchin A.;
 RA "Secretion of cyclolysin, the calmodulin-sensitive adenylate cyclase-
 RT haemolysin bifunctional protein of Bordetella pertussis.";
 RL EMO J. 7:3997-4004(1998).
 CC -1- FUNCTION: CYAD IS NECESSARY FOR TRANSPORT OF CALMODULIN-SENSITIVE
 CC ADENYLATE CYCLASE-HEMOLYSIN (CYCLOLYSIN).
 CC -1- SUBCELLULAR LOCATION: Inner membrane-bound (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE HLYD FAMILY OF SECRETION PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U00039; AB18563.1; --
 CC EMBL; AE000436; AAC76610.1; --
 CC PIR; S47807; S47807.
 CC EcGene; EG12290; y1av.
 CC InterPro; IPR006143; HLYD.
 CC Pfam; PF00529; HLYD; 1.
 CC Hypothetical protein; transmembrane, signal; Complete proteome.
 CC SIGNAL 1 15
 CC CHAIN 16 378
 CC TRANSMEM 27 47
 CC TRANSMEM 57 77
 CC SEQUENCE 378 AA; 41771 MW; 6683C1E129BC3E51 CRC64;
 SO
 Query Match 6.2%; Score 125.5; DB 1; Length 378;
 Best Local Similarity 20.2%; Pred. No. 1.1;
 Matches 65; Conservative 49; Mismatches 112; Indels 95; Gaps 11;
 QY HPQV---ALTVELPGRLESRTADYRAVGIIQKRLFGSGYVRAGQPIYQIDSSYE 109
 DB 51 HPTFKAKQKAVISIP-----VVPQVGVVIEVTDKKNTLIKKEVLFRLDPTRYQ 100
 QY 110 ANLESARAQLATAQ-----ATLAKADADLARYKPLVAABA----- 144
 DB 101 ARVRLMADIVTAIEKQKALGAELDEMANANTQAKATDKAKEXORYARSSQAKVNFPS 160
 QY 145 ---VSROEYDAAVTAKRSAAEYKAKQAQAIRK-----AGININRS 181
 DB 161 ERDIDVARQNYLQABASVKSAAEQKIQSLDSLVGEHSQIALKQAQLEAKYNLEQT 220
 QY 182 RITAPISGFIQSKYSEGTLLNAGDTYVLTIRQTNPMYVNVVQASVWMLRQIAEGK 241
 DB 221 IVRAPSDGYVTQVIRPQT-----YASLPLRPVWVPIPOQKQIVAQFQNSILR 271
 QY 242 LLADGVTAIVGIRKPD--DGYVPEKGRLLFPADPVV---NSTGOITLRAAVP----- 288
 DB 272 LAPGDDAEV---FNALPGKVF--SGKLAISPAPGAYSTGTLQTLNTPASDDGVIA 326
 QY 289 -----NDQNILMPGLYVRV 302
 DB 327 TIEDEHTDLSALPDGIYQV 347
 RESULT 38
 APRE_PSEAE STANDARD; PRT; 432 AA.
 ID APRE_PSEAE
 AC 003025;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Alkaline protease secretion protein aprE.
 GN APRE OR PA1247.
 OS Pseudomonas aeruginosa.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 CC Pseudomonadaceae; Pseudomonas.
 CC NCBI_TaxID=287;
 CC
 RN (1)
 RN SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=93051361; PubMed=1427098;
 RA Duong P., Lazdunski A., Cami B., Murgier M.;
 RT "Sequence of a cluster of genes controlling synthesis and secretion
 RT of alkaline protease in Pseudomonas aeruginosa: relationships to
 RT other secretory pathways";
 RL Gene 121:47-54(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,

RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Gabler R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
 RA Brody R.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 RT opportunistic pathogen.";
 RL Nature 406:959-964(2000).
 CC -1- FUNCTION: INVOLVED IN THE SECRETION OF ALKALINE PROTEASE.
 CC -1- SUBCELLULAR LOCATION: Inner membrane-bound (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE HLYD FAMILY OF SECRETION PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X64558; CAA45856.1; --
 CC EMBL; AE004554; AAC04636.1; --
 CC PIR; G83489; G83489.
 CC PIR; S26697; S26697.
 CC InterPro; IPR006143; HLYD.
 CC InterPro; IPR006144; HLYD_FAMILY.
 CC InterPro; IPR003997; RtxD.
 CC Pfam; PF00529; HLYD; 1.
 CC PRINTS; PR01490; RTXTOXIND.
 CC PROSITE; PS00543; HLYD_FAMILY; 1.
 CC Transmembrane; Inner membrane; Transport; Complete proteome.
 CC DOMAIN 1 14
 CC TRANSMEM 15 36
 CC TRANSMEM 37 432
 CC DOMAIN 378 379
 CC COMFLCT 378 379
 CC SEQUENCE 432 AA; 48052 MW; 4C845C65DECF15EE CRC64;
 SO
 Query Match 6.1%; Score 123; DB 1; Length 432;
 Best Local Similarity 21.6%; Pred. No. 1.8;
 Matches 79; Conservative 56; Mismatches 99; Indels 132; Gaps 17;
 QY 57 VALVELPGRLESRTADYRAVGIIQKRLFGSGYVRAGQPIYQIDSSYEANLESAR 116
 DB 42 VPAIVTIIISGGRKS-----VOHPLGCVVKHILVDDGVEAGEPLIRMEPTQARANVDSL 96
 QY 117 AQLATQATLAKADADLARYKPLVAABAASVROEYDAVTKRSAE--AGYKAAQAIRKSA 174
 DB 97 NRYANARLNOARLQA-----EYD-----GRRTLEMPAGL-AEQAPLPTL 134
 QY 175 G--ININR-----SRITAPISGFIQ---SKVSEGT-----LN-----A 204
 DB 135 GERLEIKORQLHSRQTLANELSLRANIGLRQLEGLRQTBGNORLQORLNSQLSGA 194
 QY 205 GD-----TTVLTIRQTNPMYVNVVQASVWMLRQIAEGKLLAA----- 245
 DB 195 RDLAESGYMRNQLBERQLAEVNAIRLSSSGRFGQIRQSIADQAQRIARQREBEYRKEV 254
 QY 246 DGYIA-----VGIK-PDQGVYPEKGRLLFPAD 271
 DB 255 NGOLAETQVNAARTLWELLSARAYELRHAETRAPVSGYVAGIKVFTDGGVI----- 304
 QY 272 PVNNESTGOITLRAAVVNDQNILMPGLYVRVLMDOV---AVDNAFVVPQAAVTRGAKDT 327
 DB 305 -----GPEGL-LWYIVNSDSLEVEGQALANLVDRHISGLPEVMTLPAFNQSKTPRVTGE 358
 QY 328 VMIYNA 333
 DB 359 VTMVSA 364
 RESULT 39
 YA02_RHIME STANDARD; PRT; 525 AA.
 ID YA02_RHIME

DR Pfam: PF00529; HlyD: 1.
 DR PRINTS: PRO1490; RIKTOXIND.
 KW Hypothetical protein; Transmembrane; Inner membrane; Coiled coil.
 FT TRANSMEM 9 28 POTENTIAL.
 FT DOMAIN 148 207 COILED COIL (POTENTIAL).
 FT NON TER 338 338
 SQ SEQUENCE 338 AA; 37548 MW; 5F56F234A9B42FEE CRC64;

Query Match 6.3%; Score 128; DB 1; Length 338;
 Best Local Similarity 23.0%; Pred. No. 0.68;
 Matches 73; Conservative 43; Mismatches 101; Indels 100; Gaps 13;

QY 60 TVEPLGRLESLRTADYRAQVGIIIOKRLFOEGSYVRAQGLYOIDSSTYEANLESARA-- 117
 DB 36 TVTIGNIDIRQVQAAPFDNGRLDLRV-OEGDPRVKKQQLADLPVRFQADVPKDAVW 94
 QY 118 -----QLATQAQTLAKADADLARYKPLVAEASRQEVAA 153
 DB 95 PRKSRCPACRVPVPRKSPRPGKPAQAQTLNNAETWROQALARQYVPRKOSLDNA 154
 QY 154 VTAKRSAGVYAAQAAI-----KSAGINLANR-----SRITAPI 187
 DB 155 AALAKTARANDRAQQAALTLAIKGRKEDIAAARQQLADKAGLSLARRELDRLVAP 214
 QY 188 SGFTGQSVSEGTILNAGD---TTVLATITQTPMYVNTQSASEVMKLRQIAEGKLL 243
 DB 215 DG-----VVQDRILEPDMVSPQTPVFTLALDNPVWV-----RAYLPEKAL- 255
 QY 244 AADGVIAVGIFKFDGTV-----YPERK---GRLLFADPVV--NESTGQIT-----LRA 285
 DB 256 ---GQVRLGKMK---ATISSDSFPKGSFPGKWFGISFTLAEFTKVTQTELTELTVYRV 309
 QY 286 AVPDNDITLMPGLYVVR 302
 DB 310 YACNPQHRRLRGMPVTV 326

RESULT 36
 AN36_HELPY STANDARD; PRT; 329 AA.
 ID AN36_HELPY
 AC P94851;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 36 kDa antigen.
 GN HPI488.
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
 CC Helicobacteriaceae; Helicobacter.
 CX NCBI_TaxID=210;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=26695 / ATCC 700392;
 RX MEDLINE=97394467; PubMed=9252185;
 RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
 RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
 RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
 RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
 RA McInerney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
 RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
 RA Cotton M.D., Weidman J.W., Fujii C., Bowman C., Watthey L., Wallin E.,
 RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
 RA Venter J.C.;
 RT "The complete genome sequence of the gastric pathogen Helicobacter
 pylori."
 RL Nature 388:539-547(1997).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AEO00647; AAO08523.1;
 DR EMBL: U86610; AAB81283.1;
 DR PIR: H64705; H64705.
 DR TIGR: HPI488;
 DR InterPro: IPR003997; REXD.
 DR PRINTS: PRO1490; RIKTOXIND.
 KW Antigen; Transmembrane; Complete proteome.
 FT TRANSMEM 11 31
 FT VARIANT 67 67 H -> R (IN STRAIN HP921023).
 FT VARIANT 256 256 E -> D (IN STRAIN HP921023).
 FT VARIANT 272 272 K -> R (IN STRAIN HP921023).
 FT VARIANT 274 274 T -> A (IN STRAIN HP921023).
 SQ SEQUENCE 329 AA; 36199 MW; B33A75727056DF87 CRC64;

Query Match 6.3%; Score 127.5; DB 1; Length 329;
 Best Local Similarity 22.8%; Pred. No. 0.7; Indels 67; Gaps 6;
 Matches 63; Conservative 38; Mismatches 108;

QY 63 LPGRLESLRTADYRAQVGIIIOKRLFOEGSYVRAQGLYOIDSSTYEANLESARA----- 117
 DB 38 LQGFLEA-REYVSASKVGRIEKVFVKGDHIKKGDLVFSISPELEAKLAQAEAGKAA 96
 QY 118 -----QLATQAQTLA-----KADADLAR 135
 DB 97 KALSDVEYKGRSRETTINSARDVWQAQKSAQTLAKETKRVQDDYDNGVASLQKRDENVAA 156
 QY 136 YKPLVAEASRQEVAAV-----TAKRSAGVYAAQAAIKSAGINLANSRTAPISGF 190
 DB 157 YESTKYNESAPYQTKMALGASASESKIAKAKESAPALGOVNEVESYLKQVKATAPIDG 216
 QY 191 IGQKSVSEGTILNAGDTTVLATTITQTPMYVNTQSASEVMKLRQIAEG----- 240
 DB 217 VSNVLISGGLSPKGFVVMIDLKQSMWLKISPEKYNLFYKXER-BEYITALKSKSTK 275
 QY 241 ---KLADGVIAVGIFKFDGTVPEKGRLLFADPV 273
 DB 276 FRVYKVLVWGDPAFTWKATNNSNTYDMKSYEVEAIPL 311

RESULT 37
 YIAV_ECOLI STANDARD; PRT; 378 AA.
 ID YIAV_ECOLI
 AC P37683;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein yiaV precursor..
 GN YIAV OR B3586.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; Escherichia.
 CX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=94316500; PubMed=8041620;
 RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
 RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
 RT region from 76.0 to 81.5 minutes."
 RL Nucleic Acids Res. 22:2576-2586(1994).
 CC -I- SIMILARITY: STRONG, TO E.COLI YIIB; SOME, TO P.AERUGINOSA APPE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

```

CC -1- SUBUNIT: Interacts with HDAC7. Forms a large corepressor complex
CC that contains SIN3A/B and histone deacetylases HDAC1 and HDAC2.
CC This complex associates with the thyroid (TR) and the retinoid
CC acid receptors (RAR) in the absence of ligand.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS:
CC   Name=Long;
CC   Name=Alternative splicing; Named Isoforms=2;
CC   IsoId=Q60974-1; Sequence=Displayed;
CC   Name=Short;
CC   IsoId=Q60974-2; Sequence=VSP_003411;
CC -1- TISSUE SPECIFICITY: Ubiquitous.
CC -1- DOMAIN: The N-terminal region contains repression functions that
CC are divided into three independent repression domains (RD1, RD2
CC and RD3). The C-terminal region contains the nuclear receptor-
CC interacting domain that are divided in two separate interaction
CC domains (ID1 and ID2).
CC -1- DOMAIN: The two interaction domains (ID) contain a conserved
CC sequence referred to as the corn box. This motif is required and
CC sufficient to permit binding to unliganded TR and RARs. Sequences
CC flanking the CORN BOX determine nuclear hormone receptor
CC specificity.
CC -1- SIMILARITY: Contains 1 SANT-A domain.
CC -1- SIMILARITY: Contains 1 MYB-like domain.
CC -1- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U35312; AAB17125.1; -;
DR EMBL; U22016; AAC52168.1; -;
DR TRANSFAC; T04688; -;
DR MGI; MGI:1349717; Ncor1.
DR InterPro; IPR001005; MYB_DNA_binding.
DR Pfam; PF00249; myb_DNA-binding; 2.
DR SMART; SM00717; SANT; 2.
DR PROSITE; PS50090; MYB 3; 1.
KW Nuclear protein; Transcription regulation; DNA-binding; Repressor;
KW Coiled coil; Alternative splicing.
FT DOMAIN 174 216 COILED COIL (POTENTIAL).
FT DOMAIN 254 312 INTERACTION WITH SIN3A/B (BY SIMILARITY).
FT DOMAIN 299 328 COILED COIL (POTENTIAL).
FT DNA_BIND 437 482 SANT-A (POTENTIAL).
FT DNA_BIND 501 550 MYB.
FT DOMAIN 501 550 COILED COIL (POTENTIAL).
FT DOMAIN 606 616 PRO-RICH.
FT DOMAIN 2073 2077 CORN BOX OF ID1.
FT DOMAIN 2277 2281 CORN BOX OF ID2.
FT DOMAIN 58 64 POLY-GLN.
FT DOMAIN 593 602 POLY-PRO.
FT DOMAIN 1044 1047 POLY-ALA.
FT DOMAIN 1713 1718 POLY-ALA.
FT DOMAIN 1968 1979 POLY-SER.
FT VASAPLIC 2333 Missing (in isoform Short).
FT CONFLICT 1952 1952 I -> T (IN REF. 2).
FT CONFLICT 2090 2090 A -> P (IN REF. 2).
SQ SEQUENCE 2453 AA; 270640 MW; 52208B40382F7E6A CRC64;
Query Match 6.4%; Score 130; DB 1; Length 2453;
Best Local Similarity 21.3%; Pred. No. 4.9;
Matches 106; Conservative 70; Mismatches 164; Indels 158; Gaps 26;
QY 28 GGDAAAG-----GPPAREAPPVGVVYVYVQVVALTVLPGRLSELRAD-----74
DB 1066 GGSISGPTGTYLSSHNQAYPOEAPKPSVG-----SISGLPQOESTKAAPLTYI 1116

```

```

QY 75 -----VRAQVGIIQ--KRLFOEGSVYRAGQPLYQIDSTYEANLES 114
DB 1117 KQEFSPRSQNSQPEGLLVAAQHGGVVRGTAAGVQESIR--GTPAKISVET-----ISS 1171
QY 115 ARAQL-----ATQA--TLAKADADLARYKPL-----VAAEAVR-----147
DB 1172 LRSITGTPALPQAGIPTEALVKGPIVSR--MPIEESSEPKYREBAASKGHVYEGKSGHI 1230
QY 148 QEYDAANTAR-----SAEAGVKAQAARKAGINLRSRITAPISGFI-----191
DB 1231 LSYDNIKNAAEGTRSPPTAHESMLKSEYAEVSGSIQK--GMSMBESPVASPDELICPALP 1289
QY 192 ----GQSKVSEGLTLNAGDTIVLATIRQTNMYVNVVTSASEVMKLRQIAEGK--LLAAD 246
DB 1290 RGSFPHSLKERTYLS-----GSIMGTR--ATAESFPDGLKVPQIKRESPIPAFE 1340
QY 247 GVIAVGIRPDGTVYPERKGLP-----ADPVNVEST-----GQITLRAPV 288
DB 1341 GAITKGKPYDGIITTIKEMGRSHEIPRODILTQESRKTPEVOSTRPIEGSISQGTPIK 1400
QY 289 NDQN-----ILMPGLYVRLMDQVANDAFVVPQ--QATPRGAKOTVMVNA 333
DB 1401 FDNNSGQSALIKHNVKSLITGPSKLPROMLE-----IYVENIKVERKRYED--VKA 1449
QY 334 QGGMEPREVTVAQQGQGNMIVTGLKDGKVVEGSIAGI--TGAKKVPKWEASSENQ 391
DB 1450 GEVYRHRHTSVSSGPE--VLRSTLHEAPKAQIS----PGLYDSSARRIPVQYNTISR 1503
QY 392 AAAPQSGVGTASBAKTAS 409
DB 1504 GSPMMNRITSVSSSKSAS 1521
RESULT 35
YBHG THIFE STANDARD; PRT; 338 AA.
ID YBHG THIFE
AC Q9L9D4;
DT 26-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein in asrc 5' region (ORF1) (Fragment).
OS Thibacillus ferrooxidans.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Acidithiobacillales;
OC Acidithiobacillaceae; Acidithiobacillus.
OX NCBI_TaxID=920;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33020;
RX MEDLINE=20250620; Pubmed=10788346;
RA Butcher B.G., Deane S.M., Rawlings D.E.;
RT "The chromosomal arsenic resistance genes of Thibacillus ferrooxidans
RT have an unusual arrangement and confer increased arsenic and antimony
RT resistance to Escherichia coli."
RL Appl. Environ. Microbiol. 66:1826-1833(2000).
CC -1- SUBCELLULAR LOCATION: Membrane-associated. Inner membrane
CC (Potential).
CC -1- SIMILARITY: BELONGS TO THE UPF0194 FAMILY.
CC -1- CAUTION: Comparisons with its orthologs suggests that this protein
CC may be full-length.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF173880; AAF69246.1; -;
DR HAMAP; MF_01304; -; 1;
DR InterPro; IPR006143; Hyd.
DR InterPro; IPR003997; RtxD.

```

OY 378 KKATPKWASSENQAAAFQS 397
 Db 514 LLITAEAMIAIDIPERGSPQS 533

RESULT 33

EMRK_ECOLI STANDARD; PRT; 387 AA.
 ID EMRK_ECOLI P76517; P76944; P76945;
 AC P52599; P76517; P76944; P76945;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Multidrug resistance protein K.
 GN EMRK OR B2368.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Uesumi R.;
 RL Submitted (NOV-1995) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [3]
 RP SEQUENCE FROM N.A.

RC STRAIN=K12;
 RX MEDLINE=97349980; PubMed=9205837;
 RA Yamamoto Y., Alta H., Baba T., Hayaishi K., Inada T., Isono K.,
 RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
 RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,
 RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubram S.,
 RA Tengan H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
 RA Yamagata S., Horiuchi T.;
 RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
 RT - K12 genome corresponding to 50.0-68.8 min on the linkage map and
 RT analysis of its sequence features.";
 RL DNA Res. 4:91-113(1997).
 CC -1- SUBCELLULAR LOCATION: Inner membrane-bound (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE HLYD FAMILY OF SECRETION PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

CC EMBL, D78168; BAA11236.1; ALT INIT.
 DR EMBL, AEO00325; AAC75427.1; -
 DR EMBL, D90867; BAA1239.1; ALT INIT.
 DR EMBL, D90866; BAA1231.1; -
 DR PIR, E65010; E65010.
 DR EcoGene, EG13282; emrk.
 DR InterPro, IPR005694; Emr.
 DR InterPro, IPR006143; HLYD.
 DR InterPro, IPR006144; HLYD_FAMILY.
 DR Pfam, PF00529; HLYD; 1.
 DR TIGRFAMs, TIGR00998; Baa101; 1.
 DR PROSITE, PS00543; HLYD_FAMILY, FALSE_NEG.
 KM Transport; Transmembrane; Inner membrane; Complete proteome.
 FT DOMAIN 1
 FT TRANSMEM 17 37
 FT POTENTIAL.

FT DOMAIN 38 387 PERIPLASMIC (POTENTIAL).
 SQ SEQUENCE 387 AA; 42585 MW; ABAD26F5D68FF38F CRC64;
 Query Match 6.4%; Score 133.5; DB 1; Length 387;
 Best Local Similarity 19.4%; Pred. No. 0.37;
 Matches 61; Conservative 53; Mismatches 106; Indels 95; Gaps 9;

OY 75 VRAVGGIIQKRLFEQGSVYRAQPLVQIDSTYEANLESAR----- 116
 Db 57 ISAVSGSVTVVNHKDTNYVQGDILVSLDKTATIALNKAKNLNIIVRTNKLVLQDK 116
 OY 117 ---AQATTAQTAKADADLARKPLVAEAVRQEV---DAVTKRAEAGVKAQA 169
 Db 117 QYSAEVASARIQYQSLSDYNNRVPLAKGVISKETLEHTKTLISSKALNAIQVKA 176
 OY 170 -----AIKSGININRRIPTAFISFGISQSVSEGLTNAGDPT 208
 Db 177 NKALWMTPLNRPQVVEADATKEMALKRTDISPVTGYIAQNSVQGVYVSPQS- 235
 OY 209 VLATIRQTNPMYVNTQSASEVKKLRQIAEGKLLADGVIAVGIRPDGTVPKGRRL 268
 Db 236 -LMAVVPARQMWVNNANFKETQLTDVR--IGQVNIISD-----LYGE----- 274
 OY 269 FADPVNVESTGQITLRAAVNDONILMPGLYRVYLMDOVAVDNAF-VVPOQAVTRGAKDT 327
 Db 275 -----NVVPHG---RVTGIMGTGNAFSLPQNAATGMWIKI 308
 OY 328 VMIVNAQGMREPV 342
 Db 309 VQRPVEVSIIDPKEL 323

RESULT 34

NCRL_MOUSE STANDARD; PRT; 2453 AA.

AC Q60974; Q60812;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Nuclear receptor co-repressor 1 (N-COR.) (N-COR) (Retinoid X receptor
 DE interacting protein 13) (RIP13).
 GN NCOR1 OR RXRIP13.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
 RC TISSUE=Pituitary;
 RX MEDLINE=96008539; PubMed=756614;
 RA Hoerlein A.J., Naesser A.M., Heinzel T., Torchia J., Gloss B.,
 RA Kurokawa R., Ryan A., Kamei Y., Soderstrom M., Glass C.K.,
 RA Rosenfeld M.G.;
 RT "Ligand-independent repression by the thyroid hormone receptor
 RT mediated by a nuclear receptor co-repressor.";
 RL Nature 377:397-404(1995).
 RN [2]
 RP SEQUENCE OF 1792-2453 FROM N.A. (ISOFORM LONG).
 RC TISSUE=Liver;
 RX MEDLINE=95280959; PubMed=7760852;
 RA Seol W., Choi H.S., Moore D.D.;
 RT "Isolation of proteins that interact specifically with the retinoid X
 RT receptor: two novel orphan receptors.";
 RL Mol. Endocrinol. 9:72-85(1995).
 RN [3]
 RP INTERACTION WITH HDAC7.

RX PubMed=10984530;
 RA Downes M., Ordentlich P., Kao H.-Y., Alvarez J.G.A., Evans R.M.;
 RT "Identification of a nuclear domain with deacetylase activity.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:10330-10335(2000).
 CC -1- FUNCTION: Mediates the transcriptional repression activity of some
 CC nuclear receptors by promoting chromatin condensation, thus
 CC preventing access of the basal transcription.


```

RX MEDLINE=21396508; PubMed=11481431;
RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
RA Vorhoelter F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,
RA Golding B., Puchler A.,
RT "The complete sequence of the 1,603-kb pSymb megaplasmid from the N2-
RT fixing endosymbiont Sinorhizobium meliloti."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
CC -1- SUBCELLULAR LOCATION: Membrane-associated. Inner membrane
CC (potential).
CC -1- SIMILARITY: BELONGS TO THE UPR0194 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL603645; CAC49273.1; -.
DR PIR; A95951; A95951.
DR HAMAP; MF_01304; -.
DR InterPro; IPR006143; HlyD.
DR Pfam; PF00529; HlyD; 1.
DR Hypothetical protein; Transmembrane; Inner membrane; Coiled coil;
KW Plasmid; Complete proteome.
FT TRANSMEM 5 23 POTENTIAL.
FT DOMAIN 101 208 COILED COIL (POTENTIAL).
FT DOMAIN 184 193 POLY-ALA.
SQ SEQUENCE 334 AA; 35680 MW; 9B5D1617B5E78FC3 CRC64;

Query March 6.8%; Score 136.5; DB 1; Length 334;
Best Local Similarity 22.3%; Pred. No. 0.21;
Matches 71; Conservative 46; Mismatches 109; Indels 93; Gaps 10;

QY 62 ELPGRLSLR-----TADVRA-----OVGIIQRLFGSGSYRAGPPLYQIDST 107
DB 22 DLPAFLGPARBPRQAVLYGVNDIRQVSLGFRVSGRIAEIARVDEGDSVAGDVIAKLDEP 81
QY 108 YEANIESRAQALATRAQATLAK-----ADADLA--RYKPLVA 141
DB 82 FROAVGAEEAEAEVYRATLALRLAGARAAEIAQARATHEERLAELENKALFEERAKQLRP 141
QY 142 AEAVSROEYDAAVTAKRSAGVKAQAAL-----KSAG 175
DB 142 NGTISQAEILDQANASRAADAPARSAREALVLEEGNPAEDIAAAAAQANATAKADSR 201
QY 176 INLNRSRTIADISGFIQSKVSE--GTLINAGDTVLATIRQTNPMYVNVVQSASEVMKLK 234
DB 202 ISLEDTQLAASDGI--LSRVREIGALVSPAD--IVVYLSLLEPVW-----R 246
QY 235 RQIABGKLAAAGVAVGKIPDDGVYPEKGRLLPADVYNVESTGQI-----TL 283
DB 247 AYVAEPDQGVHVPKMKVTITSDTAPDRIVEGVGFIISVAEFTPKSVTEPELRTDLYRL 306
QY 284 RAAVNDQNIIMPGLYVRV 302
DB 307 RIVIANPBGDLRQGMPTV 325

RESULT 32
CH63_RHIME
ID CH63_RHIME STANDARD; PRT; 544 AA.
AC Q93070.
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 60-kDa chaperonin 3 (Protein Cpn60 3) (groEL protein 3).
GN GROL3 OR GROL3 OR RA0064 OR SMA0124.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Plasmid pSymb (megaplasmid 1).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.

```

```

OX NCBI_TaxId=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396509; PubMed=11481432;
RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
RA Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
RA Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,
RA Kaiman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
RA Yeh K.-C., Davis R.W., Federpiet N.A., Long S.R.;
RT "Nucleotide sequence and predicted functions of the entire
RT Sinorhizobium meliloti pSymb megaplasmid."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
CC -1- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -1- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE007200; AAK64722.1; *.
DR PIR; H95269; H95269.
DR HAMAP; MF_00600; -.
DR InterPro; IPR001844; Chaperonin Cpn60.
DR InterPro; IPR002423; Cpn60/TCE-1.
DR Pfam; PF00118; Cpn60 TCP1; 1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PRINTS; PR00304; TCOMPLEXTCP1.
DR PROSITE; PS00296; CHAPERONIN60; 1.
KW Chaperone; ATP-binding; Heat shock; Multigene family; Plasmid;
KW Complete proteome.
SQ SEQUENCE 544 AA; 57521 MW; 2590163503E4479A CRC64;

Query March 6.7%; Score 134.5; DB 1; Length 544;
Best Local Similarity 21.4%; Pred. No. 0.48;
Matches 94; Conservative 69; Mismatches 190; Indels 87; Gaps 16;

QY 7 KAMRAAALAAVALVSSCGKGDAAGGAPAGREAPVAVGVTHPQVALTVLELPGR 66
DB 132 KARATKVISSSEIAQVGTIANAGDAGVEMIRAMEKVGNEGVTV----- 177
QY 67 LESRTADVAAQVGGIIQKRLFGSGSYRAGPPLYQIDSTYENLESA-----RAQLAT 121
DB 178 -EEARTADTELDV--VEGQDFRG-YL--SPFTVAERKRVLEDDPYLLIHEKVLGS 229
QY 122 AQAATLAAADADLARYKPLVAEAASVROEYDAAVTAKRSAGVKAQAALKSAGINLRS 181
DB 230 LQANLPLTEAAVQSGKRLIIISDEVEGEVLAIVNR-LNGGLKI--AAVTPFGGRK 286
QY 182 RTAPISGFIQSKVSEF-----TLNAGDTVL-----ATIR-QT 216
DB 287 AMLEDIAVLTAGQWISDEIGIKLENTVLLDMIGRARVLIKDTTIIIDSGDKASIGARV 346
QY 217 NPMYVNVVQSASEVMKLKRLQIABGKLAAAGVAVGKIPDDGVYPEKGRLLPADPVNE 276
DB 347 SQIQAQIEETASDDEKLERLAKLAGVAIVRVG-GATELEVKEKKDRI--DDALNA 402
QY 277 STGQITLRAAVNDQNIIMPGLYVRVLMDOVAV-----DNAFV-----VPQ 317
DB 403 T-----RAAV-----EESIVVGGVALLRAKSAIVGLVDNADVAVGSIYRALEAPIRQ 453
QY 318 QAVTRGAKDTVMIVNAAGQNEPREVTVAAQQGNTWIVTSGIKDGDKYVVEGSIAGITGA 377
DB 454 IADNAGVSGSIIVGKLVKVDGRDHNGQFDAQETEVDMIKAGIVPAKVVRALRDAGSIAS 513

```

RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kuhnra S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohemorrhagic *Escherichia coli*
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 CC -1- SIMILARITY: STRONG, TO E.COLI Y14V, SOME, TO P.AERUGINOSA AERE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L19044; AAC95068.1; -;
 DR EMBL; U00039; AAB18574.1; -;
 DR EMBL; AE000437; AAC76621.1; -;
 DR EMBL; AE005586; AAG58741.1; -;
 DR EMBL; AP002566; BAB37896.1; -;
 DR PIR; A86035; A86035.
 DR PIR; A98188; A98188.
 DR PIR; S47818; S47818.
 DR EcGene; EG11764; y1bH.
 DR InterPro; IPR006143; H1YD.
 DR Pfam; PF00529; H1YD; 1.
 DR Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 4 24 POTENTIAL.
 FT TRANSMEM 27 47 POTENTIAL.
 FT TRANSMEM 55 75 POTENTIAL.
 FT TRANSMEM 233 253 POTENTIAL.
 FT TRANSMEM 281 301 POTENTIAL.
 FT TRANSMEM 310 330 POTENTIAL.
 SQ SEQUENCE 378 AA; 41988 MW; 19FB46D698357EA0 CRC64;

Query Match 7.1%; Score 142.5; DB 1; Length 378;
 Best Local Similarity 22.3%; Pred. No. 0.11;
 Matches 76; Conservative 48; Mismatches 114; Indels 103; Gaps 13;

QY 42 AAPAVGVVYV-----HPQVVA--LTVELPGLESLRTADVAQVGGIIQKRLF 88
 DB 30 AATAAGVFLVSGLLIMNNHPYFTQAOKAVIALP-----IPQVTVGIVTEVTD 79
 QY 89 QEGSVRAGPPLYOIDSSTVEA-----NTESARAOLATAQAT----- 125
 DB 80 KNNOLIQGEVLFKIDPVRQARDRLADMTATHTNIKITRAQLTEQAQNTTQVSARD 139
 QY 126 -----LAKADADLAR-----YKPLVAEAVSRQEDPAAVTAK 157
 DB 140 RLFPNVQRYLKSGQAAVNPFSERDIDARQNFQAQDALVKGSAVEQAQIQSGLDSMWNGE 199
 QY 158 RSAEAGVAAQAIAKSGAGINLRKRITPISGFIQSGKVSSEGLINA-----G 205
 DB 200 QSQ---IVSLRAQLTEAKYNLEQTVIRAPNSGVYQVILRPQTYAALPLRPVWFIPDQ 256
 QY 206 DTVLATIRQNPVNVVNTQASAEVWK-LRQIAEGKLLADGVAAVGIKFDGTVYPER 264
 DB 257 KRQIVPAQRQNSLRLKPGDDAEVFNALPGQVHFGRKLTSLIPV-----PGSYQAQ 309
 QY 265 GRL--LEFADPVVNESGQITLRAAVPNDQNTIMP-GLTVRV 302
 DB 310 GVLOSITVPGTIDGLGTIELD-----PNDIDALPDGITYAQV 347

RESULT 30
 YH11 ECOLI STANDARD; PRT; 355 AA.
 AC P37626;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein yh11 precursor.

GN YH11 OR B3487.
 OS *Escherichia coli*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Escherichia*.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=X12 / MG1655;
 RX MEDLINE=94316500; PubMed=8041620;
 RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
 RT "Analysis of the *Escherichia coli* genome. V. DNA sequence of the
 RT region from 76.0 to 81.5 minutes.";
 RL Nucleic Acids Res. 22:2576-2586(1994).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U00039; AAB18463.1; -;
 DR EMBL; AE000424; AAC76512.1; -;
 DR PIR; S47707; S47707.
 DR EcGene; EG12224; yh11.
 DR InterPro; IPR006143; H1YD.
 DR InterPro; IPR003997; REXD.
 DR Pfam; PF00529; H1YD; 1.
 DR PRINTS; PRO1490; K1YTXQIND.
 DR Hypothetical protein; Signal; Complete proteome.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 355 HYPOTHETICAL PROTEIN YH11.
 SQ SEQUENCE 355 AA; 38828 MW; 6784DF1F19D5FPA CRC64;

Query Match 6.9%; Score 138.5; DB 1; Length 355;
 Best Local Similarity 26.3%; Pred. No. 0.17;
 Matches 52; Conservative 32; Mismatches 55; Indels 59; Gaps 6;

QY 65 GRLESRTADVRAVGGIIQKRLFOEGSVYRAGPPLYOIDSSTVEANLESARAOLATRA 124
 DB 41 GRIEATE-VDIASKIARIDITLVKEGFPVEGLAKMPTDRVQEQRLERIAQIKRQS 99
 QY 125 TLAKADADL-----ARYKPLVAEAVSRQEDPAV----- 151
 DB 100 AVAAQAQLLEGROSETRAQSLVNVQRAELDSVAKRTTRSRSLAQRGALISQQLDDBRA 159
 QY 152 --AAVTAKRSAEAGVKAQAIAKSGAGIN-----LNRSRITAPIS 188
 DB 160 AESARAALLESKAQVSAKAIEMARTNIIQAQTRVEAQAOTERRIADIDDSCLKAPRD 219
 QY 189 GFIGQSKVSE-GTLLNNG 205
 DB 220 GRV-QYRVAERGEVLAAG 236

RESULT 31
 YH3 RHIME STANDARD; PRT; 334 AA.
 AC Q92V44;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical membrane protein RB0873.
 GN RB0873 OR SMB21207.
 OS *Rhizobium meliloti* (Sinorhizobium meliloti).
 OC Plasmid symb (megaplasmid 2).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;

RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blotner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Valdes J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474 (1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97349980; PubMed=9205837;
 RA Yamamoto Y., Aliba H., Baba T., Hayashi K., Inada T., Isono K.,
 RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
 RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,
 RA Ogihara T., Oyama S., Saito N., Sempel J., Satoh Y., Sivaraman S.,
 RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
 RA Yamagata S., Horiiuchi T.;
 RT "Construction of a contiguous 874-kb sequence of the *Escherichia coli*
 RT - K12 genome corresponding to 50.0-68.8 min on the linkage map and
 RT analysis of its sequence features.";
 RL DNA Res. 4:91-113 (1997).
 RN [4]
 RP REVIEW.
 RX MEDLINE=94262163; PubMed=8203018;
 RA Lewis K.;
 RT "Multidrug resistance pumps in bacteria: variations on a theme.";
 RL Trends Biochem. Sci. 19:119-123 (1994).
 CC -1- FUNCTION: THE EMBL LOCUS CONFERS RESISTANCE TO SUBSTANCES OF HIGH
 CC HYDROPHOBICITY. EMBL PROBABLY PARTICIPATE IN A TRANSPORT SYSTEM
 CC TO EXTRUDE TOXINS AND DRUGS FROM THE CELL.
 CC -1- SUBCELLULAR LOCATION: Inner membrane bound.
 CC -1- SIMILARITY: BELONGS TO THE HYD FAMILY OF SECRETION PROTEINS.
 CC * This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M86657; AAA3724.1; -;
 DR EMBL: AE000353; AAC75732.1; -;
 DR EMBL: D90891; BAI6547.1; -;
 DR EMBL: D90892; BAI6552.1; -;
 DR PIR: F65048; F65048.
 DR EcGene: EG11354; emrA.
 DR InterPro: IPR005694; Emr.
 DR InterPro: IPR006143; HLYD.
 DR InterPro: IPR006144; HLYD_FAMILY.
 DR Pfam: PF00529; HLYD_1.
 DR TIGRfams: TIGR00998; 8a0101; 1.
 DR PROSITE: PS00543; HLYD_FAMILY; PLASE_NEG.
 KM Antibiotic resistance; Transport; Transmembrane; Inner membrane;
 KW Complete proteome.
 FT DOMAIN 1 25 CYTOPLASMIC (POTENTIAL).
 FT TRANSEM 26 46 POTENTIAL.
 FT DOMAIN 27 390 PERIPLASMIC (POTENTIAL).
 FT CONFLICT 62 62 I -> M (IN REF. 1).
 FT CONFLICT 138 138 K -> O (IN REF. 1).
 SQ SEQUENCE 390 AA; 42736 MW; 0FB9AB8CC0A270F9 CRC64;
 Query Match 7.5%; Score 150.5; DB 1; Length 390;
 Best Local Similarity 21.4%; Pred. No. 0.038;
 Matches 67; Conservative 53; Mismatches 108; Indels 85; Gaps 11;
 QY 75 VRAVGIGIQRLQOBSYVAGOPLYOISSTYANIESRAQLATRA-----122
 DB 64 IMSVGSVTVKVMADNTDFVEKGVLTLDPTDARQAQFEKAKTALASSVROTHQIMNSK 123

QY 123 -----OATLAKADADLARYKELVAEAVSROEYDAVTAKSAA-----162
 DB 124 OLQANIEVQKIALAKAQSDNNRRVPLGNALIGSEELGHARDATYSQAQDVAIQYNA 183
 QY 163 -----GVKAQDAIKSAGININRSRTAPISGFIQOSKVSSECTLINAQDTT 208
 DB 184 NQAMILCTKLEDQPAVOQAATEVAMWALALERTTILSPMTGYVSRVAVPGCAQISP-TTP 242
 QY 209 VLATIRGTNNMYVNVVTSASEVWKLRRQIAEGKLAD-----GVIAVGIKFPDGT 259
 DB 243 LMAVPAVIN-WWDANPKEITQIANMR--ICQPVTTITTDYGDVYTKYKV-VGLDMGTGS 298
 QY 260 VPEKGRLEFADPVVNESTQITLRAAVPNDONILMPGLVYRVLMQDVAADNAFVVPQA 319
 DB 299 AF-----SLT---PAQNAITGMWIKVQRLP-----VRIELDKQL-----EQY 333
 QY 320 VTGAKDTMIVN 332
 DB 334 PLRIGLSTLVSVN 346
 RESULT 29
 YIBH_ECOLI STANDARD; PRT; 378 AA.
 ID YIBH_ECOLI
 AC P32107;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 15-JUN-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein yibH.
 GN YIBH OR B3597 OR Z5021 OR EC84473.
 OS *Escherichia coli*, and
 OS *Escherichia coli* O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Escherichia*.
 OX NCBI_TaxID=562, 83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=93259920; PubMed=8387990;
 RA Zhao S., Sandt C.H., Feulner G., Vlazny D.A., Gray J.A., Hill C.W.;
 RT "Rns elements of *Escherichia coli* K-12: complex composites of shared
 RT and unique components that have different evolutionary histories.";
 RL J. Bacteriol. 175:2799-2808 (1993).
 RN [2]
 RP REVISION TO 205.
 RC STRAIN=K12;
 RA Hill C.W.;
 RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=94316500; PubMed=8041620;
 RA Sofia H.U., Burland V., Daniels D.L., Plunkett G. III, Blotner F.R.;
 RT "Analysis of the *Escherichia coli* genome. V. DNA sequence of the
 RT region from 76.0 to 81.5 minutes.";
 RL Nucleic Acids Res. 22:2576-2586 (1994).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grodeck E.U., Davis N.W., Lim A., Dimantanta E.T., Potamousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blatner F.R.;
 RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.";
 RL Nature 409:529-533 (2001).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,

```

RL Science 269:496-512(1995).
CC -1- FUNCTION: THE EMR LOCUS CONFERS RESISTANCE TO SUBSTANCES OF HIGH
CC HYDROPHOBICITY. EMRA PROBABLY PARTICIPATE IN A TRANSPORT SYSTEM
CC TO EXCLUDE TOXINS AND DRUGS FROM THE CELL (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Inner membrane-bound (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE HLYD FAMILY OF SECRETION PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32771; AAC22558.1; -.
DR PIR; B64101; B64101.
DR TIGR; H10898; -.
DR InterPro; IPR005694; Emr.
DR InterPro; IPR006143; HLYD.
DR InterPro; IPR006144; HLYD_FAMILY.
DR Pfam; PF00529; HLYD; 1.
DR TIGRFAMs; TIGR00998; B64101; 1.
DR PROSITE; PS00543; HLYD_FAMILY; FALSE_NEG.
DR Antibiotic resistance; Transport; Transmembrane; Inner membrane;
KW Complete proteome.
FT DOMAIN 1 25 CYTOPLASMIC (POTENTIAL).
FT TRANSEM 26 46 POTENTIAL.
FT DOMAIN 47 390 PERIPLASMIC (POTENTIAL).
SQ SEQUENCE 390 AA; 43062 MW; 42CFOAFA588BAF08 CRC64;

Query Match 7.8%; Score 156.5; DB 1; Length 390;
Best Local Similarity 21.6%; Pred. No. 0.017;
Matches 71; Conservative 51; Mismatches 115; Indels 91; Gaps 10;

QY 75 VRAQVGIGIKRLPQEGSYVAGPIVQIDSTYEANLESARQATLA----- 122
DB 64 VSSQVAGVAKINADNDKVAIGILVELDPTNAKLSFEQKSNANAVRVEQLGFTVQ 123
QY 123 -----QATLKADADLARYKPLVAEAVSRQEDYDAVTAKRSAGVKAQ----- 168
DB 124 QLSAVHANEISLQAGCNLARVOLEKMGAIKDESPHAGEAYELAKANINASKNQLAA 183
QY 169 -----AAIKSAGINLRSHRTAPISGFIQSQSVSEGTLLNADPTT 208
DB 184 NQALLRNVPLRECPQIGNAINSLKQAWMLNQRTKIRSPIDGVARRNVQGVAVGAL 243
QY 209 VLATIRGTNRYVNVVTSASVMTLRQI-----AEGTLAADVIYVIGIKFDGTVPPE 263
DB 244 MAVVSNQOMLEAFKETQLTNNRIQPVKIHFDLYGKNKEFDGVIN-GIEMGTGNAF-- 300
QY 264 KGRLLFPADVNESTQGITLRAAVPNDQNLIMPLGLVYVILMDQVAVNNAFVVPQO----- 318
DB 301 --SLL-----PSQNAGMTIKVQVRVP-----VRKLD-----POQFTETP 334
QY 319 -AVTRGAKDTVMIVNAQGM-----EPR 340
DB 335 LRIGLSATAKVRISDSGAMLRKTEPRK 362

RESULT 27
YHCO_ECOLI STANDARD; PRT; 310 AA.
AC P46482;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yhcQ.
GN YHCO OR B3241.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U18997; AAA58043.1; -.
DR EMBL; AE000403; AAC76273.1; -.
DR PIR; G65116; G65116.
DR Ecogene; EG12823; yhcQ.
DR InterPro; IPR006143; HLYD.
DR Pfam; PF00529; HLYD; 1.
DR Hypothetical protein; Transmembrane; Complete proteome.
KW TRANSEM 12 32 POTENTIAL.
FT TRANSEM 12 32 POTENTIAL.
SQ SEQUENCE 310 AA; 34774 MW; 6E0232D459DC2757 CRC64;

Query Match 7.7%; Score 154.5; DB 1; Length 310;
Best Local Similarity 25.0%; Pred. No. 0.017;
Matches 73; Conservative 47; Mismatches 111; Indels 41; Gaps 11;

QY 62 ELPRLESLTADVRA---OVGGIQRKLPQEGSYVAGPIVQIDSTYEANLESARQ 118
DB 34 ESWPTRRARSADVVAIAIPVSGILITGVNHNQVLKKGILITIDOPRIQKALEEQAD 93
QY 119 LATQAATLAKADADLARYKPLVAEAVSRQEDYDAVTAKRSAGVKAQAAIKSAGINL 178
DB 94 VAVYQ-VLAQEKQEGAGRRLRGVQMSREIIOANNVLTGLVHLQAKATBDLAKLDL 152
QY 179 NRSRITPISGFIQSQSVSEGTLLNADPTVLTIRQTNRYVNVVTSASVMTLR-- 234
DB 153 ERTVIRAPADGWTNNLVNNGEFTTRGSTAV-ALVXQ-NSFYAVAYMEETKLGVRPGYR 210
QY 235 ---RQIAEGKL--AADGVAVGIKFDGTVPPEKRLPLADPVNVESTQGITLRAVFN 289
DB 211 AEITPLGSKNVYLGTVDSV-AAGVTNAST-----RDKGMATL----- 248
QY 290 DONILMGLYVRYVILMDQVAVNNAFVVPQAVTRGAKDTVMIVNAQGMPEPR 341
DB 249 DSNLEWRLAQVRVP-RIRLDNQ---QENIWPAGTTATVVVGKQDRDESDQ 296

RESULT 28
EMRA_ECOLI STANDARD; PRT; 390 AA.
AC P27303; P77356;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Multidrug resistance protein A.
GN EMRA OR B2685.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RP SEQUENCE FROM N.A.
RX MEDLINE=93028382; PubMed=1409590;
RA Lomovskaya O., Lewis K.;
RT "Emr", an Escherichia coli locus for multidrug resistance.";
Proc. Natl. Acad. Sci. U.S.A. 89:8938-8942(1992).

```

```

Db      161 LKSNQDKLSQYRTGNREODIAQAASLEQAQAOLADQDITLLIPANGTLLTRAVE 220
Qy      198 EGTLLNMGDTVLTATIRQTNPMY-----NTVQASAEWKLKRLQIAEGTLAADGVIA 250
Db      221 PGMNMGSTVL--TLSTLRVWRAVYVDERNLSQT-----QPGRDILLTYDG--- 266
Qy      251 VGIRFDGTVYPER--GRLLFADPVVNESTGQIT-----LRAAVPDQ 291
Db      267 -----RPDKRYHGKIGFVSP-----TAEFTPKVTPEPDLRTDLYVRLRIIVDAD 311
Qy      292 NLMPLGLVYRVLMQVA 308
Db      312 DALRQGMPTVTKFNDEA 328

RESULT 25
CUSB_ECOLI
ID CUSB_ECOLI STANDARD; PRT; 407 AA.
AC P77239;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative copper efflux system protein cusb precursor.
GN CUSB OR B0574.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
CX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-VIDes J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RT Science 277:1453-1474 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,
RA Federpriel N., Hyman R., Kaiman S., Komp C., Kurdi O., Lew H., Lin D.,
RA Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kinura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizouchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiiuchi T.;
RT "A 711-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12,728,0 min region on the linkage map.";
RT DNA Res. 3:137-155 (1996).
RN [4]
RP GENE NAME.
RC STRAIN=K12 / DH5-alpha;
RX MEDLINE=20461235; PubMed=11004187;
RA Munson G.P., Lam D.L., Outten F.W., O'Halloran T.V.;
RT "Identification of a copper-responsive two-component system on the
RT chromosome of Escherichia coli K-12.";
RL J. Bacteriol. 182:5864-5871 (2000).
CC -1. FUNCTION: MAY BE A COMPONENT OF A COPPER EFFLUX SYSTEM.
CC -1. SIMILARITY: TO S. TYPHIMURUM SLB.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sb-sib.ch).
CC -----
Cc      EMBL; AB000162; AAC73675.1; -
DR      EMBL; U82598; AAB40772.1; -
DR      EMBL; D90699; BAA35208.1; -
DR      EMBL; D90700; BAA35214.1; -
DR      PIR; D64790; D64790.
DR      EcoGene; EG14235; cusb.
KM      Transport; Copper; Signal; Complete proteome.
FT      SIGNAL 1 POTENTIAL.
FT      CHAIN 29 407 PUTATIVE COPPER EFFLUX SYSTEM PROTEIN
FT      CUSB.
SQ      SEQUENCE 407 AA; 44304 MW; E949390000C61FC6 CRC64;

Query Match 8.0%; Score 162; DB 1; Length 407;
Best Local Similarity 24.1%; Pred. No. 0.0083;
Matches 84; Conservative 51; Mismatches 141; Indels 72; Gaps 13;

Qy      50 VTHHPQVALTVLPGRLSRLRTADVRAQVGGIQRKL-FOEGSYRAGOPLYQIDSTY 108
Db      99 VTGPIPLFPAOSFPANVSNEYQYAIQVQARAAGFIDKVPPLTVGDKVKGTPLDLTPDW 158
Qy      109 -EANTESARAQLATAQTLAKA-----DADLARYKPLVAEAVSROEYDAVTA 156
Db      159 VEAQSEVLLRLREGTATQTEGILERLRLAGMPEADIR--LIATQKIQ----- 205
Qy      157 KRSAAEAVKAAQAIAKAGINLNSRTTAPISGFIGSKVSEGLTAGDTTVLATIRQT 216
Db      206 -----TRFTLKAPIDGVTITAFDLRAG--NMIAKDNNVAKIQGM 241
Qy      217 NPMVYN--VTQASAEWKLKRLQIAEGKLLAADGVIAVGIFKFDGTVYPERKRLFPADPV 274
Db      242 DPVAVTAIFESTIAMLVKDSQF-----TLTVPAR-PDKTLLIRKTKLL--PGV 287
Qy      275 NESTGQITLRAVAVNDONILMPGLYVRLMDQVAVDNAFVVPQA-VTRGAKDTMTVNA 333
Db      288 DAATRTQLRLVEDNADNADLAKPGNNMVLQLN-TASEBMLLIPSGALIDTSGEGVITVDA 346
Qy      334 QGGMEPREVTVAQQGNTNIVTSGLDGDKVYVEGI---SIAGITGA 377
Db      347 DGRFVPRVAVFQASQGVTAIRSGLAEGEKVSSGFLIDSEANISGA 394

RESULT 26
EMRA_HAEMIN
ID EMRA_HAEMIN STANDARD; PRT; 390 AA.
AC P44928;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Multidrug resistance protein A homolog.
GN EMRA OR H10898.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
CX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KM20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Goodeyne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cocton M.D.,
RA Uettermann T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";

```

DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Hypothetical membrane protein ybhg.
 GN YBHG OR STY0853 OR T2070.
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=601;
 RP SEQUENCE FROM N.A.
 RC STRAIN=CT18;
 EX MEDLINE=21534947; PubMed=11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagers K.,
 RA Krogan A., Larsen T.S., Leach S., Moute S., O'Gaora P., Parry C.,
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrall B.G.;
 RA "Complete genome sequence of a multiple drug resistant Salmonella
 RT enterica serovar Typhi CT18.";
 RL Nature 413:848-852(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Ty2 / ATCC 700931;
 RX MEDLINE=22531367; PubMed=12644504;
 RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
 RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
 RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
 RT and CT18.";
 RL J. Bacteriol. 185:2330-2337(2003).
 CC -1- SUBCELLULAR LOCATION: Membrane-associated. Inner membrane
 CC (potential).
 CC -1- SIMILARITY: BELONGS TO THE UPF0194 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AL627268; CAD05265.1; -
 CC EMBL: AE016841; AAC69690.1; -
 CC DR HAMAP: MF_01304; -; 1.
 CC DR InterPro: IPR006143; HlyD.
 CC DR Pfam: PF00529; HlyD; 1.
 CC KM Hypothetical protein; Transmembrane; Inner membrane; Coiled coil;
 CC Complete proteome.
 CC TRANSMEM 5 27 POTENTIAL.
 CC DOMAIN 140 209 COILED COIL (POTENTIAL).
 CC SEQUENCE 331 AA; 36339 MW; 6D65583979B5F48 CRC64;
 FT
 SQ
 Query Match 8.3%; Score 168.5; DB 1; Length 331;
 Best Local Similarity 22.7%; Pred. No. 0.0027;
 Matches 72; Conservative 49; Mismatches 89; Indels 107; Gaps 10;
 QY 70 LRTADVAVQVGGIIQKRLFOEGSYVRAGQPLVYQIDSSITYEANLESARQAQLATAQATL-- 126
 Db 41 IRTVNISFRVGGRLASLNVDGDTIKAGVYLGELDHAPYENALMQAKAGVSVAQAQYDLM 100
 QY 127 -----AKADADLA-----RYKPLVAEAVERQEDAVTAKRSADAG 163
 Db 101 LAGYRDEEIAQAAAVRQAAYVAQNFYNRQOGLMKSRITISANDLENASSRDQAQAT 160
 QY 164 VKAAQ-----AAIKSAGINLNSRITAPISGFIGOSKVS 197
 Db 161 LKSKQDRLISQYRTGNREODIAQAASLEQAKAQIAQAOGLDQDTLLTAPAGTLLTAAVE 220
 QY 198 EGTLLNAGDTTATIRQTNPMV-----NTVQASAEVWKLRQIAEGTLAADVIA 250

Db 221 PGMNMGSTYL--TLSLRPVRWRAVYDERNLSQT-----QPGRDILLYTDG--- 266
 QY 251 VGKIFDQGVYPER---GRLLFADPVVNESIGQIT-----LRAAVPDQ 291
 Db 267 -----RPDPYHKGIFVSP-----TAETPKTVERPDLRTLVRLRIIVDDAD 311
 QY 292 NIMPGLYRVYLDQVA 308
 Db 312 DALRGWPFVTKFNDFA 328
 RESULT 24
 YBHG_SALTY STANDARD; PRT; 331 AA.
 ID YBHG_SALTY
 AC 082QPF0;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE Hypothetical membrane protein ybhg.
 GN YBHG OR STM0818.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=602;
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
 RT LT2.";
 RL Nature 413:852-856(2001).
 CC -1- SUBCELLULAR LOCATION: Membrane-associated. Inner membrane
 CC (potential).
 CC -1- SIMILARITY: BELONGS TO THE UPF0194 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AE008733; AAL19755.1; -
 CC STyGene: SG27272; ybhg.
 CC DR HAMAP: MF_01304; -; 1.
 CC DR InterPro: IPR006143; HlyD.
 CC DR Pfam: PF00529; HlyD; 1.
 CC KM Hypothetical protein; Transmembrane; Inner membrane; Coiled coil;
 CC Complete proteome.
 CC TRANSMEM 5 27 POTENTIAL.
 CC DOMAIN 140 209 COILED COIL (POTENTIAL).
 CC SEQUENCE 331 AA; 36310 MW; 6754949E5033BE83 CRC64;
 FT
 SQ
 Query Match 8.3%; Score 168.5; DB 1; Length 331;
 Best Local Similarity 22.7%; Pred. No. 0.0027;
 Matches 72; Conservative 49; Mismatches 89; Indels 107; Gaps 10;
 QY 70 LRTADVAVQVGGIIQKRLFOEGSYVRAGQPLVYQIDSSITYEANLESARQAQLATAQATL-- 126
 Db 41 IRTVNISFRVGGRLASLNVDGDTIKAGVYLGELDHAPYENALMQAKAGVSVAQAQYDLM 100
 QY 127 -----AKADADLA-----RYKPLVAEAVERQEDAVTAKRSADAG 163
 Db 101 LAGYRDEEIAQAAAVRQAAYVAQNFYNRQOGLMKSRITISANDLENASSRDQAQAT 160
 QY 164 VKAAQ-----AAIKSAGINLNSRITAPISGFIGOSKVS 197

RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome
RL corresponding to the 12.7-28.0 min region on the linkage map";
CC DNA Res. 3:137-155(1996).
CC -1- SUBCELLULAR LOCATION: Membrane-associated. Inner membrane
CC (Potential).
CC -1- SIMILARITY: BELONGS TO THE UPF0194 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; AE000181; AAC73882.1; -;
CC DR EMBL; D90717; BAA35461.1; -;
CC DR EMBL; D90716; BAA35455.1; -;
CC DR PIR; C64816; C64816;
CC DR EcGene; EG13315; ybhG.
CC DR HAMAP; MF_01304; -; 1.
CC DR InterPro; IPR006143; HlyD.
CC DR Pfam; PF00529; HlyD; 1.
CC KW Hypothetical protein; Transmembrane; Inner membrane; Coiled coil;
CC Complete proteome.
CC FT TRANSMEM 7 26 POTENTIAL.
CC FT DOMAIN 107 210 COILED COIL (POTENTIAL).
CC SQ SEQUENCE 332 AA; 36416 MW; A61DC1D96F84785E CRC64;

Query Match 8.5%; Score 171.5; DB 1; Length 332;
Best Local Similarity 23.6%; Pred. No. 0.0018;
Matches 71; Conservative 42; Mismatches 101; Indels 87; Gaps 7;

QY 70 LRTADVRAQVGIIQKLFQEGSYVRAQGYQIDSTYEANLBSARAQATLTAQATL-- 126
DB 42 IRTNLSRVRGGRVSLAVDGDALIKAGQVGLDHRKYEIALMOKAKGVSAQAQVDLM 101
QY 127 -----AKDADIA-----RYKPLVAEAVSROEYDAVTAQRSAEAG 163
DB 102 LAGYRNEEIAQAAAAVKAQQAAYDAQNFVRQOGLMWSRTISANDLENARSRDQAAT 161
QY 164 VKA-----AAAIKRSAGINRSRTAPISGFGQSVS 197
DB 162 LKSNODKLRQYRSGNREDDIAQASLEQAQAQALNLDSTLLAPSGTLLTRAWE 221
QY 198 EGTLLNAGDTVLATIRQTNPMYVNTQASAEVVKLRQIAEGKLLADGVIAVGKFD 257
DB 222 PGTYLNEGCT--VFTVSLTRPVW-----RAYVDERNLDQAQPKRVLLVTDG 267
QY 258 GTVYPEKGRLLFADPVNVESTGQIT-----LRPAVPNDQIIMPLGLYR 301
DB 268 RPDKRYHQIGFVSP-----TAEFPKTVETPDRTDLVYRLRYVTDADALRQGMFVT 322
QY 302 V 302
DB 323 V 323

RESULT 22
YDHU ECOLI STANDARD; PRT; 285 AA.
AC P76185; P76893;
DT 15-JUL-1998 (Rel. 36, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein ydhU.
GN YDHU OR B1644.
OS *Escherichia coli*.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; *Escherichia*.

OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of *Escherichia coli* K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kaai H., Kashimoto K., Kimura S., Kitakawa M., Kikugawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T.,
RA Sempel G., Seki Y., Sivasubraman S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the *Escherichia coli* K-12 genome
RL corresponding to the 28.0-40.1 min region on the linkage map";
CC DNA Res. 3:363-377(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; AE000259; AAC74716.1; ALT_INIT.
CC DR EMBL; D90807; BAA15404.1; -;
CC DR EcGene; EG13944; ydhU.
CC DR InterPro; IPR006143; HlyD.
CC DR Pfam; PF00529; HlyD; 1.
CC KW Hypothetical protein; Transmembrane; Complete proteome.
CC FT TRANSMEM 6 26 POTENTIAL.
CC FT CONFLICT 284 285 GQ -> VNNRIYVWLAQFALVQCHAGAMALCTQYCHVS
CC FT GADGCLLFKSG (IN REF. 2).
CC SQ SEQUENCE 285 AA; 31496 MW; FC3032D13764A869 CRC64;

Query Match 8.4%; Score 169; DB 1; Length 285;
Best Local Similarity 25.4%; Pred. No. 0.0021;
Matches 71; Conservative 43; Mismatches 105; Indels 60; Gaps 9;

QY 52 VHPQVALTVLPGRLSLRTADVRAQVGIIQKLFQEGSYVRAQGYQIDSTYEAN 111
DB 40 IBAQVGSITQVSGRIVELNIKD-----NQLVNAAGDLLLTIDKTPQIA 83
QY 112 LESARAQATRAQATLAAADADLARYKPLVAEAVSROEYDAVTAQRSAGVQAQDAI 171
DB 84 ELNAQAQALAAQSGSLAANNEANRRRL--SQNFISAEELDTANLVNAPAKASVDAQAATL 142
QY 172 KSAGINLRSRTAPISGFGQSVSEGTLLNAGDTVLATIRQTNPMYVNTQASAEV 231
DB 143 KQAWQALAQTEIRAPVSGWVNTLTTRIGDYADTG-----KPLFALVDSHSRYVI 191
QY 232 -----KLRLQIAEGK-----LAAAG-----YVAGIKRFDGTVPEKGRLLFADPVN 275
DB 192 GYFEETKL-RHIREGAPAQITLVSDNKTLOGHVSISGRAIYDQSVESDSSI-----PDVK 246
QY 276 ESTGQITLRAAAPPNDQIIMPLGLYRVLMQVAVDNAFV 314
DB 247 PNPVFWLAQRP-----YRPALDKVPDQVTLV 274

RESULT 23
YBHG SALTI STANDARD; PRT; 331 AA.
AC Q82879;
DT 15-JUL-1998 (Rel. 36, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein ybhG.
GN YBHG OR B1644.
OS *Escherichia coli*.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; *Escherichia*.


```

QY 154 VTAKSAAGVKA-----OAAIKSAGININRSRTAPI 187
DQ 152 RTANQARANQAQKADKLAQFLSGNRPGEIAQAEANLAQTAELAQQLNADOTLLAPS 211
QY 188 SGFIGQSKVSEGLTLNAGDTTLATIROTNPMYVNTOSASEVMKLRQIAEGKLLAADG 247
DQ 212 AGTLTAVREPGTILASANT--VFTVSLTDPVWV-----RAYVSERHGAQAP 257
QY 248 VIANIGVEDGTVPPEK--GRLLPADVNVESGQIT-----LRANVP 288
DQ 258 GSEVEV-FTDGR--PDKPYHKGIVGSP-----TAEFTKPTVEPDLRTDLVRLRIIT 309
QY 289 NDQNIIMPGLYVRVLMQ 306
DQ 310 DADSLRGMPVTVTRFPQ 327

RESULT 18
YJCR_ECOLI STANDARD; PRT; 343 AA.
AC P32716;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein yjcr.
GN yjcr OR B4082.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=94089392; PubMed=8265357;
RA Blatter F.R., Burland V.D., Plunkett G. III, Sofia H.J.,
RA Daniels D.L.;
RT "Analysis of the Escherichia coli genome. IV. DNA sequence of the
RT region from 89.2 to 92.8 minutes."
RL Nucleic Acids Res. 21:5408-5417(1993).
CC -1- SUBCELLULAR LOCATION: Inner membrane-bound (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE HLYD FAMILY OF SECRETION PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U00006; AAC43176.1; -.
DR EMBL; AE000481; AAD13465.1; -.
DR PIR; A65217; A65217.
DR EcoGene; EG11954; yjcr.
DR InterPro; IPR005694; Emt.
DR InterPro; IPR006143; HlyD.
DR Pfam; PF00529; HlyD; 1.
DR TIGRFAMs; TIGR00998; Bae0101; 1.
KW Hypothetical protein; Transmembrane; Inner membrane; Transport;
KW Complete proteome.
FT DOMAIN 1 12 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 13 33 POTENTIAL.
FT DOMAIN 34 343 PERIPLASMIC (POTENTIAL).
SQ SEQUENCE 343 AA; 36908 MW; C599B870C475FFE CR64;

Query Match 8.8%; Score 178.5; DB 1; Length 343;
Best Local Similarity 23.9%; Pred. No. 0.00073;
Matches 87; Conservative 46; Mismatches 116; Indels 115; Gaps 12;

QY 6 FKARAAALAAVAVLVSSCGKGGDAAGGQAPAGREAPVGVVTVHPQVATLVELPG 65
DQ 13 FPAALLVAL-ALVALVF-----VIVRVDSAPSTNDAYASADITDIVPEVSG 57

```

```

QY 66 RLSELRADVRAQGVGIIQKRLPQEGSVYRAGPLVQIDSEYANLESARAQLA----- 120
DQ 58 RIVELAVATDQA-----VKQGDLLPRIDRPYBANLAKAASLAALDKQ 101
QY 121 -----TAQTLAKADA-----DLARYPLVAEAVSROEYDAATYAK 157
DQ 102 IMLTORSVDAQORGAOSVNATVEKARAAKQADTDLRTPELKEGFVSAEDVDRASTAQ 161
QY 158 RSAEAGVKA-----AQAIRKAG-----INLANSRTAPISGFIGQ 193
DQ 162 RAADPADANAVLLQKQSAASVSGVDALVQAPAAVEADIALTKHLEMATVRAPPDGVIS 221
QY 194 SKVSEG-----TLNAGDTTLATIROTNPMVY-VNTOSASEVMKLRQIAEGKL 242
DQ 222 LKTSVGGFASAMRPFLIDTRHMYVIANFREDDKNIRSGTATIRIMSDSGKTEGKV 281
QY 243 LAADGVAVGIRKDDGTVPYKGRLLPADVNVESGQITLRAAVPDQNIIMPGLYVRV 302
DQ 282 -----DSIGYGLPDDG-----GLVIGLGPVKVRSINWVAVQORFP-----VKI 320
QY 303 LMDQ 306
DQ 321 MVDK 324

RESULT 19
YBHG_ECOLI6 STANDARD; PRT; 332 AA.
AC Q8RJN6;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hypothetical membrane protein ybhg.
GN YBHG OR C0878.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:HI / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Raeko D., Buckles E.L., Ikon S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Moley H.L.T., Domeneberg M.S., Blatter P.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli."
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
CC -1- SUBCELLULAR LOCATION: Membrane-associated. Inner membrane
CC (potential).
CC -1- SIMILARITY: BELONGS TO THE UPF0194 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AF016757; AAN79351.1; -.
DR HAMAP; MF_01304; -.
DR InterPro; IPR006143; HlyD.
DR Pfam; PF00529; HlyD; 1.
KW Hypothetical protein; Transmembrane; Inner membrane; Coiled coil;
KW Complete proteome.
FT TRANSMEM 7 26 Potential.
FT DOMAIN 108 211 Coiled coil (Potential).
SQ SEQUENCE 332 AA; 36402 MW; 75DFB761CEB9885BA CR64;

Query Match 8.6%; Score 173.5; DB 1; Length 332;
Best Local Similarity 23.7%; Pred. No. 0.0014;

```

DT 28-FEB-2003 (Rel. 41, last sequence update)
 DT 28-FEB-2003 (Rel. 41, last annotation update)
 DE Nucleation protein nolP.
 GN NOLP OR RA0481 OR SMA0876.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Plasmid pSymA (megaplasmid 1).
 CC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 CC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
 CX NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AK631;
 RX MEDLINE=91360053; PubMed=1909418;
 RA Baev N., Endre G., Petrovics G., Banfalvi Z., Kondorosi A.;
 RT "Six nodulation genes of nod box locus 4 in Rhizobium meliloti are
 RT involved in nodulation signal production: nodM codes for
 RT D-glucosamine synthetase.";
 RL Mol. Gen. Genet. 228:113-124 (1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=21396509; PubMed=11481432;
 RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
 RA Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
 RA Gurjal M., Hong A., Hutzar L., Hyman R.W., Kahn D., Kahn M.L.,
 RA Kaiman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
 RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
 RT "Nucleotide sequence and predicted functions of the entire
 RT Sinorhizobium meliloti pSymA megaplasmid.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888 (2001).
 CC -1- FUNCTION: INVOLVED IN THE PRODUCTION OF MEDICAGO-SPECIFIC
 CC NODULATION SIGNAL MOLECULE.
 CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FIVE
 CC FRAMESHIFTS IN POSITIONS 78, 83, 119, 308 AND 329.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X58632; CAA41486.1; ALT_FRME.
 DR EMBL; AE007238; AAK65139.1; -.
 DR PIR; A95322; A95322.
 DR InterPro; IPR006143; HLYD.
 DR Pfam; PF00529; HLYD.1.
 KM Plasmid, Nodulation; Complete proteome.
 FT CONFLICT 9 L -> V (IN REF. 1).
 FT CONFLICT 43 K -> E (IN REF. 1).
 FT CONFLICT 82 E -> D (IN REF. 1).
 FT CONFLICT 160 OL -> HV (IN REF. 1).
 FT CONFLICT 189 A -> R (IN REF. 1).
 FT CONFLICT 237 T -> M (IN REF. 1).
 FT CONFLICT 301 A -> T (IN REF. 1).
 FT CONFLICT 358 V -> A (IN REF. 1).
 FT CONFLICT 363 II -> YV (IN REF. 1).
 SQ SEQUENCE 367 AA; 39031 MW; 67793C426F21E9BF CRC64;
 Query Match 9.1%; Score 184; DB 1; Length 367;
 Best Local Similarity 23.0%; Pred. No. 0.00037;
 Matches 88; Conservative 62; Mismatches 182; Indels 50; Gaps 12;
 QY 11 AALAAVALVLSGCGKSGDAAGGAPAGREAPAVGVTVHQTALVTELPGRLESL 70
 DB 13 AFATTLAVTTVLSS-GPAIGQVKGHSP-----IELAKADVSTAVRQDVANEVTVGSLTPI 67
 QY 71 RTADVRAQVGGIIQKRLFOEGSYVRAGOPLYOIDSSTYEANLESARQLATAQTATLAKD 130
 DB 68 RRTITRSVSGSTIELFVQIGDVNAGDLVPERGALLESVATCRKAADALSAQETLAE 127
 QY 131 ADLARKPLVAAEVSRQEVDAVTAKRSAAGVKAQAALIKSAGININRSRTITAPISGF 190

DB 128 AVLEERNTRLGERGASASATRLAALADVLRLAQLRSQAQVSDERSLSHAENVAERFGCV 187
 QY 191 IGQSVSEGLTLNAGDTTVALTIRQTNPMYNTVQSASEWKLK-RQIAEGLLAADGVI 249
 DB 188 IAAVSVEGQVPL-NTQMLTVELNRLVEDAGVPTSRPLTLKQSV-----LTVEG-- 240
 QY 250 AVGIKFDGTVYPERGKLLPADPVNVESTQITLRAAVPNDQNLMPGLVYRVLMQVAV 309
 DB 241 -----PGRTFSGEVARI-----SPTADAGSRARVFIADVNDDEGLLRGGMFT---IGDLRV 289
 QY 310 D---NAVAVPQAVTRAKQDTWIVNAQGMREPREVVAQOQGNW-----ITYSGLKDG 361
 DB 290 DDQKDVIALPASIRHDADGFVLKYRAGVLQRRPVG-----GRSWSDRDLVOVSGVSEG 345
 QY 362 D-----KVVVEGI 369
 DB 346 DVIVTAPLPLVNVNTPIIIGI 367
 RESULT 17
 Y873 YERPE STANDARD; PRT; 328 AA.
 ID Y873 YERPE
 AC Q93AT7;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, last sequence update)
 DT 28-FEB-2003 (Rel. 41, last annotation update)
 DE Hypothetical membrane protein Ecs0873.
 GN Ecs0873.
 OS Yersinia pestis.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; Yersinia.
 CX NCBI_TaxID=632;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Yokohama / Biovar Antiqua;
 RA Radnedge L., Agron P.G., Worsham P.L., Andersen G.L.;
 RT "Genome plasticity in Yersinia pestis.";
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Membrane-associated. Inner membrane
 CC (Potential).
 CC
 CC -1- SIMILARITY: BELONGS TO THE UPF0194 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AF426171; AAL27384.1; -.
 DR HAMAP; MF_01304; - 1.
 DR InterPro; IPR006143; HLYD.
 DR Pfam; PF00529; HLYD.1.
 KM Hypothetical protein; Transmembrane; Inner membrane; Coiled coil.
 FT TRANSMEM 7
 FT DOMAIN 80 112 POTENTIAL.
 FT DOMAIN 139 212 COILED COIL (POTENTIAL).
 SQ SEQUENCE 328 AA; 35950 MW; C9B905A0C4D1D36A9 CRC64;
 Query Match 9.1%; Score 183; DB 1; Length 328;
 Best Local Similarity 24.8%; Pred. No. 0.00038;
 Matches 79; Conservative 48; Mismatches 97; Indels 94; Gaps 11;
 QY 60 TVELPRLSLRTADVRAQVGGIIQKRLFOEGSYVRAGOPLYOIDSSTY-----EANL 112
 DB 33 SLTLYGNDV-IRTVNIGFRVAGRLASLAVDEGDIDHPGQTLGKLDGPYLNALKQAVNV 91
 QY 113 ESRAPL-----ATAQTLAKADADLARKPLVAAEVSRQEVDA 153
 DB 92 QSAQALALLKAGYREEETIAQVSEVAQROAAFDVADNPLKQOGLMASKAVSANLEENA 151

OS Alkaligenes eutrophus (Ralstonia eutropha).
OG Plasmid pMOL28.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=510;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=CH34;
RX MEDLINE=93139046; PubMed=8380802;
RA Liesegang H., Lemke K., Siddiqui R.A., Schlegel H.-G.;
RT "Characterization of the inducible nickel and cobalt resistance
determinant chr from pMOL28 of Alkaligenes eutrophus CH34.";
RL J. Bacteriol. 175:767-778(1993).
CC -1- FUNCTION: NICKEL AND COBALT RESISTANCE PROTEINS CNRA, CNRB, CNRC
CNRH AND CNRR MAY BE INVOLVED IN THE REGULATION OF CNR.
CC -1- SIMILARITY: SOME, TO A. EUTROPHUS C2GB.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M91650; AAA21969.1; -
DR PIR; F47056; F47056.
DR InterPro; IPR006143; HYD.
DR InterPro; IPR005695; Membrane_fus2.
DR InterPro; IPR006311; Tat.
DR Pfam; PF00529; HYD; 1.
DR TIGRFAMs; TIGR00999; 8a0102; 1.
DR TIGRFAMs; TIGR01409; Tat signal seq; 1.
KW Plasmid; Transport; Nickel; Cobalt;
SQ SEQUENCE 395 AA; 40082 MW; 075C2165B5B8A09F CRC64;
Query Match 9.6%; Score 194; DB 1; Length 395;
Best Local Similarity 25.5%; Pred. No. 0.0001;
Matches 100; Conservative 55; Mismatches 167; Indels 70; Gaps 14;
QY 14 LAAVALVLSGCGKGD--AAAGGAPAGREAP-----AVY----- 46
DB 13 MIAGVAAVAAVGAHLPVSEKSPASTQAPBAQKPOSAPVPGKKEKIPATVYLAAN 72
QY 47 -----VGVTVHPQTALVTELPGRLSRTDPAVRAQVGIIOKRLFOEGSVYRA 96
DB 73 IAVEPVASAAVGTETLAPATA--ALPG-----SEAVTSRAAGAVQVRRLGDVYKA 124
QY 97 GQPLVQIDSTYEANLESARQALTAQTLAKADADLARVYKPLVAEAESVROEYDAVTA 156
DB 125 GDVIALVDSF--EAGMAAEKRVNAQKADLARKTYE--REASLFOQGVTPROEMAAKKA 180
QY 157 KRSAGVKAQAQAIKSAGI-NLNRS-RITAPISGFIGQSKVSEGLTLNAGDTTVLATIR 214
DB 181 LDVAQAELRAATYAQSAHLASDGRSVAVSPVIAKITAQSVTLGAFV--APQAELFRVA 238
QY 215 QTNEMVYVNTQSASEVMKLRQIAEGKLADGVIAVGIKRPDDGVYPERKRLLPADPV 274
DB 239 GTGAVQVAAVTAADTSRIAG-SEATLLANG-----SFLSARVAQVPTV 284
QY 275 NES--TGQITLRAAVPNDONILMPGLYVRLMDQVAVNAPVFPQAVTRGAKDTVMIVN 332
DB 285 TGSARVATVVVPAQPTDRLVYEGSVQR-LRTAVADAALASVBDANQNDGRDVLFR 343
QY 333 AAGGMEPREVTVAAQOQGTNMTVTSGLKDGDKV 364
DB 344 TQEGFRPMPVLYVTRSGGSAQILSGVQAGEQV 375

DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Putative membrane fusion protein silB precursor.
OS SilB.
GN Salmonella typhimurium.
OC Plasmid pMG101.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=CH34;
RX MEDLINE=99128056; PubMed=9930866;
RA Gupta A., Matsui K., Lo J.-F., Silver S.;
RT "Molecular basis for resistance to silver cations in Salmonella.";
RL Nat. Med. 5:183-188(1999).
CC -1- FUNCTION: COMPONENT OF THE SIL CATION-EFFLUX SYSTEM THAT CONFERES
RESISTANCE TO SILVER. MAY BE PART OF A THREE-COMPONENT
CATION/PROTON ANTI-PORTER.
CC -1- SIMILARITY: TO E. COLI CUSB.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF067954; AAD11748.1; -
DR Transport; Antiport; Signal; Plasmid.
KW SIGNAL 1
FT CHAIN 29 430 PUTATIVE MEMBRANE FUSION PROTEIN SILB.
SQ SEQUENCE 430 AA; 47719 MW; A3269C6D5F7F8095 CRC64;
Query Match 9.3%; Score 187; DB 1; Length 430;
Best Local Similarity 23.3%; Pred. No. 0.0003;
Matches 97; Conservative 70; Mismatches 168; Indels 82; Gaps 17;
QY 19 ALVLSGCGKGD-----DAAAGGAPAGREAPVGVTVHPQTALVTELPLRLE 68
DB 10 ALIISLILAGLISVTAMQYLNSSQKTVPBAQKPEKKV-LFWYDPKPKPTKPKGKSP 68
QY 69 SL-----RTADYRA--QVGI-----IOKRLFOEGSVYRAGQPLVQIDSTYEANLES 114
DB 69 FMDMDLVPKYADESGKSSGIRIDPTQVQ-NLGKTKQKXTRGLANT--SQITPAVSY 124
QY 115 ARAQLATAQTLAKADADLARVYKPLVAEAESVROEYDAVTAKSAAE----- 162
DB 125 NEVQFVIVQ--ARSDGFVEKVPPLTIGDHVKGTPLIDITIPBWEAQSEFLLSGTG 181
QY 163 ---GVKAQAALISAGI--NLNRS-----ITAPISGFIGQSKVSEGLTLNAGD 206
DB 182 TPTQIKGVLERLRLRAGPREDIQRLRSTRITQRTFKAPIDGVITTFDIDRTG--NMISK 239
QY 207 TVTIAATIRQTNPMVNV--VTQSASEVMKLRQIAEGKLADGVIAAGIKRPDDGVYPER 264
DB 240 DKVVAQIQGMDPVIAAIPESIAVYLKQDSQPE-----ISVP-AYPDITFIVEK 288
QY 265 GRLLFADPVNVNESTQITLRAAVPNDONILMPGLYVRLMDQVAVNAPVFPQAVTRGA 324
DB 289 WNLV---PSVDQTRTLQVRLQVTNKRDEFLKPGNNAVLKLNTQSQEWMLLPSQAVIDTQK 345
QY 325 KQIVMYVNAAGGMEPREVTVAAQOQGTNMTVTSGLKDGDKVBEI-----SIAGITGA 377
DB 346 FQVITVDEGKFPVQKQIHVLHESQOOSGIGSLNEDGVVVSGLFLIDSEANITGA 402

RESULT 16
NOLF_RHIME
ID NOLF_RHIME STANDARD; PRT; 367 AA.
AC P25196;
DT 01-MAY-1992 (Rel. 22, Created)

RA Kunio T., Kusano T., Oyaizu H., Senoo K., Kanazawa S., Matsumoto S.;
 RT "Cloning and sequence analysis of czc genes in Alcaligenes sp. strain
 RL CT14.";
 CC Biosci. Biotechnol. Biochem. 60:699-704(1996).
 CC -1- FUNCTION: CZCA AND CZCB TOGETHER WOULD ACT IN ZN(2+) EFFLUX
 CC NEARLY AS EFFECTIVELY AS THE COMPLETE CAC EFFLUX SYSTEM (CZCABC).
 CC THE CACB PROTEIN IS THOUGHT TO PUMP Zn(2+) CATIONS TO THE CZCA
 CC TRANSPORT PROTEIN.
 CC -1- SIMILARITY: SOME, TO A. EUTROPHUS CNRB.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D67044; BA01060.1; -.
 CC PIR: J04699; J04699.
 CC InterPro: IPR006143; HLYD.
 CC InterPro: IPR005695; Membrane_fus2.
 CC Pfam: PF00529; HLYD; 1.
 CC TIGRfams: TIGR00999; 8a0102; 1.
 CC Plasmid; Transport; Zinc; Cobalt; Cadmium resistance.
 CC KW SEQUENCE 520 AA; 54574 MW; AF7000ABA955C224 CRC64;
 SO
 Query Match 11.7%; Score 236; DB 1; Length 520;
 Best Local Similarity 25.7%; Pred. No. 4.8e-07;
 Matches 102; Conservative 57; Mismatches 168; Indels 70; Gaps 12;
 QY 9 MRAALAAVALVSSCGKGDAAQGGPA-----GREAPAPV----- 46
 Db 126 VRATGESQALKFVVS-----GDALSSQGPVAPHPVDTANTVTLPGSSSLAVRLSKSEG 180
 QY 47 -----VGVV--TVHPQYALTYELRG--RLSELRTADPRAQVGGIIOKRLPFG 91
 Db 181 KIELTADQALKTGVVQVQAGSAGKAVQAGVQFGEIRFNEDEKTAHVPRLAGVESVPANIG 240
 QY 92 SYVRAGQPLVQIDSSYEANLESARQATLATAQATLAKADADLARYKPLVAEAVSROEYD 151
 Db 241 QOVKKQGLVLAIVAST-----GLSDQSELLAAOKRLDLARVYDEKLEKMEKISADEYL 296
 QY 152 AAVTAKSAEAGVKAQAQAIKISAGIN-----LNSRRTAPISGFIGOSKVSSEGLTLNAGD 206
 Db 297 SARNAIQEAQISVQNAQOKLTALIGASNSSTALNRYELAPAPAGMIVEGHISLGEAV--AD 354
 QY 207 TTVLALTRQTPMVTYVNTQSAEYWKLRQIAEGKL--AADGYIANGIKFDDGTIVPEK 264
 Db 355 NANFTLSDLSVAAEFVSAKADVERVR--IGEKASINSASSDVYA-----DGTV----- 402
 QY 265 GRLLFADPVNVESTQGITLRAAVPNDONILMPGLVYRVLMDQVAVDNAFVVPQQAIVTGA 324
 Db 403 ----SYVSLGLSEQRTAKARVTLTNPQAMRPGLVTVDFGADVEVPAVKTEAVDDVN 459
 QY 325 KDTVMIVNAQGMREPVTVAAQOQGTNMIIVTSLKDG 361
 Db 460 GESVVFVAVGGGFVPQPVKVGRTNGKVIIEVGLKPG 496
 RESULT 13
 NCB ALCX STANDARD; PRT; 397 AA.
 ID NCB ALCX 044585;
 AC 044585;
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DE Nickel-cobalt-cadmium resistance protein ncb.
 GN NCB.
 OS Alcaligenes xylosoxydans xylosoxydans (Achromobacter xylosoxydans).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; Achromobacter.

OX NCB TaxID=515;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=31A;
 RX MEDLINE=95050278; PubMed=7961470;
 RA Schmidt T., Schlegel H.G.;
 RT "Combined nickel-cobalt-cadmium resistance encoded by the ncc locus
 RL of Alcaligenes xylosoxydans 31A.";
 CC J. Bacteriol. 176:7045-7054(1994).
 CC -1- FUNCTION: COMPONENT OF THE NCC CATION-EFFLUX SYSTEM THAT CONFERS
 CC RESISTANCE TO NICKEL, COBALT AND CADMIUM.
 CC -1- SIMILARITY: HIGH, TO A. EUTROPHUS CNRB AND SOME, TO ALCALIGENES
 CC CZCB.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L31363; AAA65105.1; -.
 CC PIR: I39579; I39579.
 CC InterPro: IPR006143; HLYD.
 CC InterPro: IPR005695; Membrane_fus2.
 CC Pfam: PF00529; HLYD; 1.
 CC TIGRfams: TIGR00999; 8a0102; 1.
 CC Plasmid; Nickel; Cobalt; Cadmium; Transport.
 CC KW SEQUENCE 397 AA; 39933 MW; C6D857CE31499688 CRC64;
 SO
 Query Match 10.8%; Score 217.5; DB 1; Length 397;
 Best Local Similarity 28.1%; Pred. No. 4.4e-06;
 Matches 113; Conservative 51; Mismatches 149; Indels 89; Gaps 20;
 QY 12 AALAAVALVSSCGKGDAAQ--GAPAGRE----APPVGVTVHPQYALTYELPGR 66
 Db 16 AGVAAAALV-----GFGARGLGSPSGAEVSKLAAPEKAAASA--PAAPAEVRIFG- 67
 QY 67 LESRTADVAAQ--VGSI-----IOKRLFQSGSY 93
 Db 68 -EYLAANINAVEPVSAGGVGSVLAPASVAAPGSEVAVISRAAGAVLRIOKRL--GDA 123
 QY 94 VRAGQPLVQIDSSYEANLESARQATLATAQATLAKADADLARYKPLVAEAVSROEYDAA 153
 Db 124 VRADVLAIVDSR--EAAAMAERKVAQARADLARKYE--RESSLQOQGTTPQEMESA 179
 QY 154 VTAKSAEAGVKAQAQAIKISAGINLN--RS--RTAPISGFIGOSKVSSEGLTLNAGDTVLA 211
 Db 180 RIALDVQAQAEVQRAATVAQAQAKVSSDGRSVAVSPPIAGRTAQSVTLG-----A 228
 QY 212 TIRQTNMYVNVVQSAEYWKLRQIAEGKLAD--GVIANGIKFDGTIV-----PEKG 265
 Db 229 YVAPQAELEFVAGSAGAVV-----EAYVTADTSRIAAG--SPATVLANGAPLAG 277
 QY 266 RLFLFADPVNVESTQGITLRAAVPNDON--ILMPGLVYRVLMDQVAVDNAFVVPQQAIVTGA 322
 Db 278 RVQAVITFTVGSARAATV--VTPVDANSGLIIVGQVVR--LHKRAADANAMSPEDAVON 335
 QY 323 GAKDTVMIVNAQGMREPVTVAAQOQGTNMIIVTSLKDGDKV 364
 Db 336 LDGRDVVFVFTQOQFRPKSVLVGSRSGGVAQIISLVGVRGEV 377
 RESULT 14
 NCB ALCEU STANDARD; PRT; 395 AA.
 ID NCB ALCEU P37973;
 AC P37973;
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DE Nickel and cobalt resistance protein cnrb.
 GN NCB.

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB005269; AAG55260.1; ALT_INIT.
DR EMBL; AP002553; BAB34387.1; ALT_INIT.
DR EMBL; AE015112; AAN42471.1; ALT_INIT.
DR InterPro; IPR006143; H1YD.
DR Pfam; PF00529; H1YD; 1.
DR Transport; Inner membrane; Signal; Antibiotic resistance;
KM Complete proteome.
FT SIGNAL 1
FT CHAIN 32 371 MACROLIDE-SPECIFIC EFFLUX PROTEIN MACA.
SQ SEQUENCE 371 AA; 40639 MW; 25F7D3CB1A2D080F CRC64;

Query Match 13.2%; Score 267.5; DB 1; Length 371;
Best Local Similarity 28.1%; Pred. No. 4.6e-09;
Matches 96; Conservative 53; Mismatches 166; Indels 27; Gaps 7;

QY 42 APAPVGVVTVHPQTVALTELPGRLSRTADYRAVGIIQKRLFOEGSYVRAGOPLY 101
DQ 31 APVFTYQTLIVRPEDLQGSVATGKLDLRKVDYGAQVSGQKTLVAIGDKVKDDLLG 90
QY 102 QIDS-----STYEANLESARAQATQAATLAKADADLARYKPLVAEAVSRQEDYDAV 154
DQ 91 VIDEQAENQIKVEYEATIMELRAQROQAELKLARVYSHQRLAQTOAVSLQDLDPTAA 150
QY 155 T-AKRSAG-----VKAQAALKSAGININRSRITRPAISFGIGQSVSG-TLLNAGD 206
DQ 151 TEMVKAQAQIGTIDAQIKRNASLDTAKTNLDYRIYAPMAQGEVQTITTLGGQTVIAAQ 210
QY 207 TTVLATIRQTNPMYVNTQSASEVMKLRQIAEGKLLAADGVIAVGIKFDGTVYPEKGR 266
DQ 211 APNLTILADMSITLVKQAVSEADYHILKPGQKMFYIGDPLTRYEGIKDVLTPERK-- 268
QY 267 LLEFADPVVESTGQITLRAAVPNDONILMPGLYYRVLMDOYAVDNAFVPOQAVTRGAKD 326
DQ 269 -----VNDALFYFA-RFEVFPNPGILRLDPTAQVHILQLTVDVGNVLTILPLSALGDVPGD 320
QY 327 TWMIYV--AQQGMEPREYTVAAQOGGTNMTVITSGIKDGKVV 366
DQ 321 NRYVKLLRNGETREREVTIGARNDTVEIVKGLGAEAGDEVVI 362

RESULT 11
CZCB ALCEU STANDARD; PRT; 520 AA.
ID CZCB ALCEU
AC P13510;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cobalt-zinc-cadmium resistance protein czcb (Cation efflux system
DE protein czcb).
GN CZCB.
OS Alkaligenes eutrophus (Ralstonia eutropha).
OC Plasmid pMO130.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=510;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CH34;
RX MEDLINE=90017477; PubMed=2678100;
RA Nies D.H., Nies A., Chu L., Silver S.;
RT "Expression and nucleotide sequence of a plasmid-determined divalent
RT cation efflux system from Alkaligenes eutrophus";
RL Proc. Natl. Acad. Sci. U.S.A. 86:7351-7355(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CH34;
RA van der Leije D., Schwuchow T., Wuerztz S., Schwidetzky U.,
RA Baeyens W., Scheel P.O., Nies D.H.;
RX Submitted (JUN-1996) to the EMBL/Genbank/DBJ databases.

CC -I- FUNCTION: CZCA AND CZCB TOGETHER WOULD ACT IN ZN(2+) EFFLUX
CC NEARLY AS EFFECTIVELY AS THE COMPLETE CZC EFFLUX SYSTEM (CZCABC).
CC THE CZCB PROTEIN IS THOUGHT TO FUNNEL ZN(2+) CATIONS TO THE CZCA
CC TRANSPORT PROTEIN.
CC -I- INDUCTION: BY CADMIUM, COPPER AND ZINC.
CC -I- SIMILARITY: SOME, TO A. EUTROPHUS CNRB.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X98451; CAA67083.1; -.
DR PIR; B33830; B33830.
DR InterPro; IPR006143; H1YD.
DR InterPro; IPR005695; Membrane_fus2.
DR Pfam; PF00529; H1YD; 1.
DR TIGRPFams; TIGR00999; 8a0102; 1.
KM Plasmid; Transport; Zinc; Cobalt; Cadmium resistance.
SQ SEQUENCE 520 AA; 54587 MW; FC3D347F87A7A79 CRC64;

Query Match 11.9%; Score 241; DB 1; Length 520;
Best Local Similarity 25.9%; Pred. No. 2.4e-07;
Matches 103; Conservative 56; Mismatches 168; Indels 70; Gaps 12;

QY 9 MRPAALAAVALVLSGCKGDAQGGQPA-----GREAPAV----- 46
DQ 126 VRATGESQALKPFVVS-----GDALIESQOPVAEPHVPDVTANVTLPGSSSPLAVRLSKREG 180
QY 47 -----VGVV--TVHPQTVALTELPG--RLESLRTADYRAVGIIQKRLFOEG 91
DQ 181 KIELTDQAKTGGVTVQTAGSANYQAGVQFPGEIRFREDKTAHVPRPLAGVSVPNNTG 240
QY 92 SYVAGOPLYQIDSTYEANLESARAQATQAATLAKADADLARYKPLVAEAVSRQEDY 151
DQ 241 QYKKGQVLAIVAST-----QLSDQRSLLAAQRLDLARTYREKRLMOKISABEDYL 296
QY 152 AAVTAKRSAGYKAAQAALKSAGIN-----INRSRTADISGFIGQSKVSEGLLNAGD 206
DQ 297 SARMLQEAQISVONAQOKLTAIGASNSSTALNRVELRAPFAGIVKHSLSGAIV--AD 354
QY 207 TTVLATIRQTNPMYVNTQSASEVMKLRQIAEGKLL--AADGVIANGIFDGTYYPEK 264
DQ 355 NANVFTLISDSSVAAEFVSAKOVERYR--IGKASINSASDVKA-----DGTV---- 402
QY 265 GRLLFADPVVESTGQITLRAAVPNDONILMPGLYYRVLMDOYAVDNAFVPOQAVTRGCA 324
DQ 403 ---SYVGSILGEGQTRAKAVTILTNPMAMRPGLFVYDFGADVVPVAVKTEAVQDVN 459
QY 325 KDTVMYVNAQGGMEPREYTVAAQOGGTNMTVITSGIKDG 361
DQ 460 GESVVFVAQGVGPQPVKVRRTGKVIIEVGLKPG 496

RESULT 12
CZCB ALCEU STANDARD; PRT; 520 AA.
ID CZCB ALCEU
AC P94176;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cation efflux system protein czcb.
GN CZCB.
OS Alkaligenes sp. (strain CT14).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alkaligenaceae; Alkaligenes.
OX NCBI_TaxID=512;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96219090; PubMed=8829543;

OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [3]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RC STRAIN=K12 / M3104;
 RX MEDLINE=21429237; PubMed=11544226;
 RA Kobayashi N., Nishino K., Yamaguchi A.;
 RT "Novel macrolide-specific ABC-type efflux transporter in Escherichia coli";
 RL J. Bacteriol. 183:5639-5644(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=9742617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12";
 RL Science 277:1453-1474(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajinara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horuchi T.;
 RT "A 718-Kb DNA sequence of the Escherichia coli K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map";
 RL DNA Res. 3:137-155(1996).
 CC -1- FUNCTION: Efflux transporter for macrolide antibiotics.
 CC -1- SUBUNIT: Interacts with macb.
 CC -1- SUBCELLULAR LOCATION: Inner membrane-associated.
 CC -1- SIMILARITY: BELONGS TO THE ACRA/ACRE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AB071145; BAB64541.1; ALT_INIT.
 DR EMBL; AB000189; AAC73965.1; ALT_INIT.
 DR EMBL; D90725; BAA35597.1; -.
 DR EcoGene; EG13694; maca.
 DR InterPro; IPR006143; HLYD.
 DR Pfam; PF00529; HLYD; 1.
 KM Transport: Inner membrane; Signal; Antibiotic resistance;
 KW Complete proteome.
 FT SIGNAL 1 31
 FT CHAIN 32 371 MACROLIDE-SPECIFIC EFFLUX PROTEIN MACA.
 SQ SEQUENCE 371 AA; 40624 MW; 8BPF287DA03B92AE9 CRC64;
 Query Match 13.3%; Score 269.5; DB 1; Length 371;
 Best Local Similarity 28.1%; Pred. No. 3.5e-09;
 Matches 96; Conservative 54; Mismatches 165; Indels 27; Gaps 7;
 QY 42 APAPVGVVTHPOTVALTELPGRLESLRTADYRAOVGGIIQRFLPQEGSYVAAQGPLY 101
 DB 31 APPTTYTLIVRPDLDQSVLATKDLARKVDVGAOVSGOLKTLVAIGDKYKKDQLLG 90
 QY 102 QIDS-----STYANLIESARAQATLAQATLAKADADLARYKPLVAEAVNSRQEDYAAV 154
 DB 91 VIDPEGAENQIKVEATLTMELRAQQAABELKLAARYTSQGQQLATQKAVSQGDDLTAA 150
 QY 155 T--AKRAEAG-----VKAQAALAKSGINLNSRTIAPISGFIGSGVSEB--TLNAGD 206
 DB 151 TEMAVKQAOIGTIDAOIKRQASLDTKAKTLDYTRIIVAPMAGEVQTITTLQGGTVIAAQ 210

QY 207 TTVLATTIQTGPNMYVNVNTQASAEVWKLRRQIAEGLLAADGVIAVGKFDGTVYPEGR 266
 DB 211 APNILLTADMSAMLVKQVSEADVIHLKPGQKAMFTVLCBPLTRFEGQIKDVLPTPERK-- 268
 QY 267 LFPADPVNVESTGQITLRAVPNDQNLMPGLVYRVLMDOVAANDNAVVPQQAATRKAD 326
 DB 269 -----VNDAIFFYA-RFEVPPNPNGLRLDMDTAQVHIQLTDVKNVLTIPLSALGDVPGD 320
 QY 327 TVMIVN--AAGMEPREVTAAGQGTWMTVTSGLKDGDKVVV 366
 DB 321 NRYVKLLRNGETREREVTTIGARNDIVELVKGLRAGDEVVI 362
 RESULT 10
 MACA_EC057
 ID MACA_EC057 STANDARD; PRT; 371 AA.
 AC P68410;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Macrolide-specific efflux protein maca precursor.
 GN MACA OR Z1115 OR ECS0964 OR SF0838.
 OS Escherichia coli O157:H7, and
 OS Shigella flexneri.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=83334, 623;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=E.coli; STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grodeck E.J., Davis N.W., Lim A., Dimataria E.T., Potamouis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
 RL Nature 409:529-533(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=E.coli; STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kunara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 RT O157:H7 and genomic comparison with a laboratory strain K-12";
 RL DNA Res. 8:11-22(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES=Shigella flexneri; STRAIN=301 / Serotype 2a;
 RX MEDLINE=22722406; PubMed=12384590;
 RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
 RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Ou D., Dong J.,
 RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
 RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
 RA Yu J.;
 RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
 RT through comparison with genomes of Escherichia coli K12 and O157";
 RL Nucleic Acids Res. 30:4432-4441(2002).
 CC -1- FUNCTION: Efflux transporter for macrolide antibiotics (By
 CC similarity).
 CC -1- SUBUNIT: Interacts with macb (By similarity).
 CC -1- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE ACRA/ACRE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC -1- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE ACRA/ACRE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AJ414148; CAC90193.1; -
 CC EMBL: AE013884; AAM86365.1; -
 CC PIR: AF0166; AF0166.
 CC InterPro: IPR006143; H1YD.
 CC Pfam: PF00529; H1YD; 1.
 CC Transport: Inner membrane; Signal; Antibiotic resistance;
 CC Complete proteome.
 CC SIGNAL 1 26 POTENTIAL.
 CC CHAIN 27 371 PROBABLE MACROLIDE-SPECIFIC EFFLUX
 CC PROTEIN MACA.
 CC
 CC SEQUENCE 371 AA; 40186 MW; 13173F46A8696E3E CRC64;
 SO
 Query Match 14.7%; Score 297; DB 1; Length 371;
 Best Local Similarity 29.8%; Pred. No. 8.4e-11;
 Matches 112; Conservative 63; Mismatches 151; Indels 50; Gaps 12;
 QY 16 AANVALVSSCGKGDAAOGOPAGREAPVGVVTHPQVAL-----TVELPGLLE 68
 DB 13 AAIIVLLT-----GGFFIARHLMAP-----VPVNVIVKVRHDLQONVALATGKLD 58
 QY SLRTADVAAVGGIIOKRLFOEGSYVRAGOPLYOIDS-----STEANLESARAO LAT 121
 DB 59 AVRKVDVAQVSGLEKLYVEIGHVKGGLAMIDPQQAQNIKEVEATITQDLMAQRIQ 118
 QY 122 AQAATLAKADADLARKPLVAEAVEROEYDAVT--AKRSAEAG-----VKAQAATIKSA 174
 DB 119 AKAEHLHATVTLGRQONLAKIQVRSROLDQAVTDLAVKNAKVGITIDAKAKASLDTA 178
 QY 175 GINLNRSITAPISGFIQSKVSEB-TLLNAGDTIVLATIQTNPMYNVNTQASAEVWKL 233
 DB 179 KINLDYTOISAPMGDDVQITTLQOQTVIAAQAPNITLLADMSTMLVQAQVSEADVINTL 238
 QY 234 RROI-AEGKTLAADGVIAVGIKFDGTVYPEKGRLLFPADPVNVESTQGITLRAVPNDQN 292
 DB 239 KPGKASTTVIGDEKRRSGV-LKDILPTPK-----VNDAI-FYSARFEVPPDR 287
 QY 293 ILMGLVYRVLMDOVAADNAFVPPQAV--TRGAKDTVMIVNAQGMREPRETVAAQOQGT 350
 DB 288 LLRLQMTAQAQSIQLANVDQAIVPLAALGDELGSNRYQVTVLKGKKEKEKREVTIGIRNVV 347
 QY 351 NWIVTSGIKDGDKVVV 366
 DB 348 DAOVISGLSVEGDVIV 363
 RESULT 8
 Y894_HAEIN STANDARD; PRT; 382 AA.
 ID Y894_HAEIN
 AC Q57500;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical protein H10894.
 GN H10894.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Haemophilus.
 OX NCI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Rd / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;

RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips R., Spillig T., Hedblom E., Cotton M.D.,
 RA Utecherack T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Frazer C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
 RT Rd.";
 RL Science 269:496-512(1995).
 CC -1- SIMILARITY: BELONGS TO THE ACRA/ACRE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U32771; AAC22554.1; -
 CC TIGR: H10894; -
 CC Pfam: PF00529; H1YD; 1.
 CC Hypothetical protein; Transmembrane; Complete proteome.
 CC TRANSMEM 1 21 POTENTIAL.
 CC SEQUENCE 382 AA; 41409 MW; 7974AA9D53F0BB5E CRC64;
 SO
 Query Match 14.2%; Score 286; DB 1; Length 382;
 Best Local Similarity 25.2%; Pred. No. 3.9e-10;
 Matches 105; Conservative 66; Mismatches 148; Indels 98; Gaps 14;
 QY 6 FKARAAALAAVALVSSCGKGDAAOGOPAGREAPVGVVTHPQVALTVELPGL 65
 DB 18 FNMIRGVNISRALA-----GMP--ESSPVT-ALEVQPREMTVINTTG 58
 QY 66 RLSELRADVAAVGGIIOKRLFOEGSYVRAGOPLYOIDSSTEANLESARAO LAT 125
 DB 59 LVRRNQAGMSTQAGAVSQVLVONGNVKKGEVLVLDSDSVEGANIQAQAQASALRQT 118
 QY 126 LAKADADLARKPLVAEAVEROEYDAVTAKRSAGEVKAQAATIKSAGINLNRSRTA 185
 DB 119 -----YORVGLHNSNAVSROEMDNAKAAVDAQVASIESLKAIE-----RRKIVA 164
 QY 186 PISGFIQSKVSEBETLNAAGDTIVLATIQTNPMYNVNTQASAEVWKL 245
 DB 165 PFDEKAGIVKINVGQYVNVG--TEIVREVDITSSKVPFALSQNDLKLH---IGORVTA 218
 QY 246 DGVIAVGIKFDGTVYPEKGRLLFPADPVNVESTQGITLRAAV--PNQNIILMGLVYVLM 304
 DB 219 TTDAKLETF-----SARITALEPAINSTGLVAVQATFDEDEHKLKLSGMSRLRI 270
 QY 305 DOVAVDNAFVPPQAVT-----RGAKDTVMIVN 332
 DB 271 ALPTETNQVVVPPQVAISYNNYGELAVLLEPLESEKMGSGNEKLDRLYRAKQITVFTKD 330
 QY 333 AOGMEREPTVNAQOQGTNMTIVSGLDGDKVVEGISAGITAKKVTTPREMASS 389
 DB 331 ROG-----VYALQDQNE-----VKGDXKIITGQO--GIGNSLV--EWIKD 369
 RESULT 9
 MACA_ECOLI STANDARD; PRT; 371 AA.
 ID MACA_ECOLI
 AC P75830;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Macrolide-specific efflux protein maca precursor.
 GN MACA OR B0878.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

QY 205 GDTVLTATIRGTNPMYVNTQSAEWMKLRQIAEGKLLAADGVIAVGKIPDGTIVPEK 264
 Db 194 NQADSLVTVQDLPYVLTQSVQDFLRKKEVASGQIKQVQSGTPYDNLNENKRYQOT 253
 QY 265 GRLTFADPVNESTQITLRAAVPNDONILMPGLVYRVLMQVAVDNAFVVPQAVTRGA 324
 Db 254 GTLKSDPTVDETTSVTLRAIFRPNGDILPQMVTYALVDEGSRQNVLVQDEVTNNA 313
 QY 325 --KDTMTIVNAQGMPEPREVVAQOQGTNMTVTSGLKDGKRVVVGIS--IAGITGAKV 380
 Db 314 QGKATATILDKDDVQVQLREIEASKAIGDQWVVTSGLGQGDRIYVIGLQIRIRGIARAIS 373
 QY 381 TPKEWASSENQ 391
 Db 374 SSGENASTESK 384

RESULT 6
 YEGM_ECOLI STANDARD; PRT; 415 AA.
 ID YEGM_ECOLI STANDARD; PRT; 415 AA.
 AC P76397;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein yegm precursor.
 GN YEGM OR E2074.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474 (1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97251358; PubMed=9097040;
 RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
 RA Kasai H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T.,
 RA Mizobuchi K., Mori H., Mori T., Motomura K., Nakade S., Nakamura Y.,
 RA Nishimoto H., Nishio Y., Oshima T., Saito N., Sampei G., Seki Y.,
 RA Sivasubramanian S., Tagami H., Takeda J., Takemoto K., Wada C.,
 RA Yamamoto Y., Horiuchi T.;
 RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
 corresponding to the 40.1-50.0 min region on the linkage map.";
 RL DNA Res. 3:379-392 (1996).
 CC -1- SIMILARITY: BELONGS TO THE ACRA/ACRE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL: AE000297; AAC75135.1; ALT INIT.
 DR EMBL: D90845; BAA15928.1; ALT INIT.
 DR EcoGene; EG14056; yegm.
 DR InterPro; IPR006143; HLYD.
 DR Pfam; PF00529; HLYD; 1.
 KM Hypothetical protein; Signal; Complete proteome.
 FT SIGNAL 1 21
 FT CHAIN 22 415 HYPOTHETICAL PROTEIN YEGM.
 FT SEQUENCE 415 AA; 44464 MW; 26240DBBDEBIASF CRC64;

Query Match 21.4%; Score 432; DB 1; Length 415;
 Best Local Similarity 32.2%; Pred. No. 1,le-18;
 Matches 119; Conservative 63; Mismatches 168; Indels 20; Gaps 6;
 QY 34 GGAPAGRAPAPVGVVTHPQVALTVELPGRLESRTADVPAQVGGIQLKLPQEGSY 93
 Db 49 GGRKMSGRLAPVQATVAVQAVPRVITGTTTAANTVTVRSVDCQLALHFQEQO 108
 QY 94 VRAQPLVQIDSTYEANLESAPQALTAQATLAKADADLARYPLVAEAVSRQEDYAA 153
 Db 109 VKAGDPLAEIDPQGFVKYALAAQAGQLAKDKATLANARDLARVQGLAKTNLVSGELDAQ 168
 QY 154 VTAKSAEAGVKAQAQAIAKSGINLSRITAPISGFIQSKVSEGTILNAGDTTVLATI 213
 Db 169 QALVSETEGITKADAEASVAAQQLDWSRITAPADGVRGLKQDVGVNQISSGDTTGIVVI 228
 QY 214 ROTNPMYVNTQSAEWMKLRQIAEGKLLAADGVIAVGI-KPDDGTIVPEKRLFPADP 272
 Db 229 TQTHPIDLVFTLPESDIATVVOAQAKGPLVEAMDRTNKSLSEGT-----LISLDN 281
 QY 273 VVNESTQITLRAAVPNDONILMPGLVY--RVLMQVAVDNAFVVPQAVTRGAK-DTVM 329
 Db 282 QIDATGTIKKAFNNQDDALFPNQVYNAMLVD--TEQNAVVIPTALQMGNGHFVW 339
 QY 330 IVNAQGMPEPREVVAQOQGTNMTVTSGLKDGKRVVVGI-----SIAGITGAKVTP 382
 Db 340 VLNSENKVSRLVPPGIQDSQKVVIRAGISAGDVRVTDGIDRLTEGAKVEVVAQSATTP 399
 QY 383 KEWASSENQ 392
 Db 400 EKAATREYA 409

RESULT 7
 MACA_YERPE STANDARD; PRT; 371 AA.
 ID MACA_YERPE STANDARD; PRT; 371 AA.
 AC P58411;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable macrolide-specific efflux protein maca precursor.
 GN MACA OR YP01364 OR Y2814.
 OS Yersinia pestis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Yersinia.
 OX NCBI_TaxID=632;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CO-92 / Biovar Orientalis;
 RX MEDLINE=21470413; PubMed=11586360;
 RA Parkhill J., Wren B.W., Thomson N.R., Baker S., Bentley S.D., Brooks K.,
 RA Parentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
 RA Fellwell S., Hamlin N., Holroyd S., Jagsels K., Kariyasev A.V.,
 RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrett B.G.;
 RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
 RL Nature 413:523-527 (2001).
 CC -1- SUBUNIT: Interacts with macB (By similarity).
 CC
 DR EMBL: AE000297; AAC75135.1; ALT INIT.
 DR EMBL: D90845; BAA15928.1; ALT INIT.
 DR EcoGene; EG14056; yegm.
 DR InterPro; IPR006143; HLYD.
 DR Pfam; PF00529; HLYD; 1.
 KM Hypothetical protein; Signal; Complete proteome.
 FT SIGNAL 1 21
 FT CHAIN 22 415 HYPOTHETICAL PROTEIN YEGM.
 FT SEQUENCE 415 AA; 44464 MW; 26240DBBDEBIASF CRC64;


```

RT Escherichia coli.":
RL Mol. Gen. Genet. 230:230-240(1991).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE OF 1-96 FROM N.A.
RC STRAIN=K12;
RA Klein J.R., Henrich B., Plapp R.;
RT "Molecular cloning of the envC gene of Escherichia coli.";
RL Curr. Microbiol. 21:341-347(1990).
CC -1- FUNCTION: MAY AFFECT SPECIFIC MEMBRANE FUNCTIONS, SUCH AS SEPTUM
CC FORMATION DURING CELL DIVISION, AND CELL MEMBRANE PERMEABILITY.
CC -1- SUBCELLULAR LOCATION: Attached to the inner membrane by a lipid
CC anchor (potential).
CC -1- SIMILARITY: BELONGS TO THE ACRA/ACRE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M96848; AAA02931.1; -
DR EMBL; X57948; CAA41016.1; -
DR EMBL; U18997; AAB58069.1; -
DR EMBL; AE000405; AAC76297.1; -
DR PIR; C65119; C65119.
DR EcGene; EG10266; acrc.
DR InterPro; IPR006143; HLYD.
DR Pfam; PF00529; HLYD; 1.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
DR Cell division; Signal; Inner membrane; Lipoprotein; Complete proteome.
KW SIGNAL
FT CHAIN 1 23
FT LIPID 24 385
FT CONFLICT 191 212
FT CONFLICT 313 314
FT SEQUENCE 385 AA; 41317 MW; EDBF82C853CF42A9 CRC64;
SQ
Query Match 37.2%; Score 750.5; DB 1; Length 385;
Best Local Similarity 42.1%; Pred. No. 1.7e-37;
Matches 168; Conservative 74; Mismatches 122; Indels 35; Gaps 6;
OY 17 AVALVLSGCGKGAAGGPGAGREAPAVVGVTVH-POTVALTV-ELFGRLSLRTA 73
D 17 SAALLAGNDKGEEBAHGEPO-----VTVHVKTAPELVKTELGRNNAVYIA 65
OY 74 DVRAOVGGIIQRLFOEGSYVPAGOPLYOIDSSTEYANLESARAQALTAQATLAKADL 133
D 66 EVRPVSGIVINRNFTSGSDVQAGSLYQIDPATQANYDSAKGELASAAAIAHLTV 125
OY 134 ARYPRVAAEAVSQEYDAATVAKRSAGVKAQAALKSAGININRSRTAPISGFIQ 193
D 126 KRIYPLVGTKTIISQEDIDQALADAFQADAAVIAATVESAIRINATYKVTAPISGRIGK 185
OY 194 SKVEGTLNAGDTTVALTRQTNPMYVNVTSASEWKLRRQIAEGKLADAVIAGVI 253
D 186 STVEGALVTNGQTTELATVQQLDPIYDVTSQNSDFRLKQSVAGLHKNMSTNEL 245
OY 254 KFDGCTVPEKGRLLFADPPVNESTGQTLTAAAVNDONILMPLGYAVLMDQVAVDNF 313
D 246 VMENGQTYPLKGLQIFSDVTVDSTGSLTLRAVFPNPQHTLLPGMFARIDEQVDAI 305

```

```

OY 314 VVPOQAVTRGAKD--TVMIVNAQGMREPVYTAQOQGTWIVTSGLKDGKVVVEGISI 371
D 306 LIPOGVSRTRPGATVTLVNDKSGQVEARPVVASQALGDKWLISEGLKSGDQIVVSGL-- 363
OY 372 AGITGAKKVTPEKMASSENOAAAPQSGVQTASEKTKTSE 410
D 364 -----OKARPG-----QVATVDTDPADTASK 385

RESULT 5
YHIU ECOI STANDARD; PRT; 385 AA.
ID YHIU ECOI
AC P37636;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical lipoprotein yhiu precursor.
GN YHIU OR B3513.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=94316500; PubMed=8041620;
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
RT region from 76.0 to 81.5 minutes.";
RL Nucleic Acids Res. 22:2576-2586(1994).
CC -1- SUBCELLULAR LOCATION: Attached to the inner membrane by a lipid
CC anchor (potential).
CC -1- SIMILARITY: BELONGS TO THE ACRA/ACRE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; U00039; AAB18489.1; -
DR EMBL; AE000427; AAC76538.1; -
DR PIR; S47733; S47733.
DR EcGene; EG12240; yhiu.
DR InterPro; IPR006143; HLYD.
DR Pfam; PF00529; HLYD; 1.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
DR Hypothetical protein; Signal; Inner membrane; Lipoprotein;
KW Complete proteome.
FT SIGNAL 1 20
FT CHAIN 21 385
FT LIPID 21 21
FT SEQUENCE 385 AA; 41190 MW; 2C825B6CDE15C70F CRC64;
SQ
Query Match 32.5%; Score 656; DB 1; Length 385;
Best Local Similarity 36.5%; Pred. No. 6.4e-32;
Matches 143; Conservative 71; Mismatches 153; Indels 4; Gaps 2;
OY 25 CGKGGDAAGGPGAGREAPAVVGVTVH-POTVALTV-ELFGRLSLRTADVRAOVGGIIQ 84
D 14 CGAMLTACDDKSAENAAAMPPEVGVTVLSFGSVNVVSELGRITVPYVAELRPVGGII 73
OY 85 KRLFOEGSYVPAGOPLYOIDSSTEYANLESARAQALTAQATLAKADLARYKPLVAAEA 144
D 74 KRNIFEDKYNQGGSLYQIDPAPLOAEINLSAKGSLAALSTASVARTTFNRQASLLKTNV 133
OY 145 VSRQEVDAATVAKRSAGVKAQAALKSAGININRSRTAPISGFIQSGSVSSGTLNNA 204
D 134 VSRQDYDTARTQLNEAEANVTAAVAEQAATINLOAVNVTSPLTGVSGKSSVTGALVTA 193

```



```

Db 61 VELPGRLESLRTADVRAQVGGIIORLFLQEGSVYRACQPLVQIDISTYEAGLESARAOLA 120
Qy 121 TAOATLAKADADLARYKPLVAEAVSROEYDAATVAKRSAGVYAAQAATIKSAGINNR 180
Db 121 TAOATLAKADADLARYKPLVADALSKOEYDAATVAKRSAGVYAAQAATIKSAGINNR 180
Qy 181 SRTAPISGFIQSGSVSEGTLLMAGDTVLATIRQTNMYNVNVSASEWKLRRQIAEG 240
Db 181 SRTAPISGFIQSGSVSEGTLLMAGDTVLATIRQTNMYNVNVSASEWKLRRQIAEG 240
Qy 241 KLLAADGVIAVGKIFDDGTVPYPERKRLIFADPVNVESTGQITLRAAVNDONILMPGLYV 300
Db 241 KLLAADGVIAVGKIFDDGTVPYPERKRLIFADPVNVESTGQITLRAAVNDONILMPGLYV 300
Qy 301 RVLMDQVAVDAFVPPQAAVTRGAKDTMTVINAQSGMPEPVYTAQOOGCTMTVTSGLKD 360
Db 301 RVLMDQVAVDAFVPPQAAVTRGAKDTMTVINAQSGMPEPVYTAQOOGCTMTVTSGLKD 360
Qy 361 GDKVVEGISIAGITGAKKVTPEKVASSENQAAAPOSQVOTASEAKTASEAE 412
Db 361 GDKVVEGISIAGITGAKKVTPEKVASSENQAAAPOSQVOTASEAKTASEAE 412

RESULT 2
ACRA_ECOLI STANDARD: PRT: 397 AA.
ID ACRA_ECOLI STANDARD: PRT: 397 AA.
AC P31223.
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Acriflavine resistance protein A precursor.
GN ACRA OR MTCR OR LIR OR B0463 OR Z0578 OR ECG0516.
OS Escherichia coli. and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334;
RN [11]
RP SEQUENCE FROM N.A.
RC STRAIN:K12.
RA Xu J., Bertrand K.P.;
RT "Nucleotide sequence of the acraB operon from Escherichia coli.";
RL Submitted (MAY-1993) to the EMBL/Genbank/DBJ databases.
RN [12]
RP SEQUENCE FROM N.A.
RC STRAIN:K12 / W4573;
RX MEDLINE=94012493; PubMed=8407802;
RA Ma D., Cook D.N., Alberti M., Pon N.G., Nikaido H., Hearst J.E.;
RT "Molecular cloning and characterization of acra and acrE genes of
RT Escherichia coli.";
RL J. Bacteriol. 175:6299-6313 (1993).
RN [13]
RP SEQUENCE FROM N.A.
RC STRAIN:K12 / MG1655;
RX MEDLINE=97426117; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Verdes J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474 (1997).
RN [14]
RP SEQUENCE FROM N.A.
RC STRAIN:K12.
RA Roberts D., Allen E., Araujo R., Aparicio A., Chung E., Davis K.,
RA Duncan M., Federspiel N., Hyman R., Kalman S., Komp C., Kuri O.,
RA Lew H., Lin D., Namath A., Oetner P., Schramm S., Davis R.W.;
RL Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.
RN [15]
RP SEQUENCE FROM N.A.
RC STRAIN:O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,

```

```

RA Postaf G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoukis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533 (2001).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN:O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasekawa C., Ogasawara N., Yasunaga T.,
RA Kubara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22 (2001).
RN [7]
RP CHARACTERIZATION.
RX MEDLINE=95379493; PubMed=7651136;
RA Ma D., Cook D.N., Alberti M., Pon N.G., Nikaido H., Hearst J.E.;
RT "Genes acra and acrB encode a stress-induced efflux system of
RT Escherichia coli.";
RL Mol. Microbiol. 16:45-55 (1995).
RN [8]
RP PROCESSING, AND INTERACTION WITH ACRA.
RX MEDLINE=20381028; PubMed=10920254;
RA Kawabe T., Fujihira E., Yamaguchi A.;
RT "Molecular construction of a multidrug exporter system, AcraB:
RT molecular interaction between Acra and AcrB, and cleavage of the
RT N-terminal signal sequence of Acra.";
RL J. Biochem. 128:195-200 (2000).
CC -I- FUNCTION: AcraB is a drug efflux protein with a broad substrate
CC specificity.
CC -I- SUBUNIT: Interacts with acrB.
CC -I- SUBCELLULAR LOCATION: Attached to the inner membrane by a lipid
CC anchor.
CC -I- SIMILARITY: BELONGS TO THE ACRA/ACRE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC
DR EMBL; M94248; AAA23410.1; -
DR EMBL; U00734; AAA67134.1; -
DR EMBL; AE000152; AAC73565.1; -
DR EMBL; U82664; AAB40217.1; -
DR EMBL; AB005225; AAG54812.1; -
DR EMBL; AP002551; BAB33939.1; -
DR PIR; A36938; A36938.
DR PIR; D90693; D90693.
DR PIR; H85543; H85543.
DR ECoGene; EG11703; acra.
DR InterPro; IPR006143; HlyD.
DR Pfam; PF00529; HlyD.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Transport; Inner membrane; Signal; Antibiotic resistance;
KW Lipoprotein; Complete proteome.
FT SIGNAL 1 24
FT CHAIN 25 397
FT LIPID 25 25
SQ SEQUENCE 397 AA; 42196 MW; 5B81DD5B2C80A077 CRC64;
Query Match 37.8%; Score 762.5; DB 1; Length 397;
Best local similarity 42.2%; Pred. No. 3.5e-38;
Matches 168; Conservative 76; Mismatches 129; Indels 25; Gaps 6;
11 AAALAAVALVLTSSCGKGGDAAGCGPAGREAPAVVGVTVHPQVALTVELPGRLESL 70

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2003, 13:56:01; Search time 17 Seconds

(without alignments)
1139.706 Million cell updates/sec

Title: US-09-889-756A-2

Perfect score: 2019
Sequence: 1 MAFYAFKAMRAALAAVAL.....AAPQGVQTAASEATASEAE 412

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1938	96.0	412	1 MTRC_NEIGO	P43505 neisseria g
2	762.5	37.8	397	1 ACRA_ECOLI	P31223 escherichia
3	753	37.3	383	1 MEXA_PSEAE	P52477 pseudomonas
4	750.5	37.2	385	1 ACRA_ECOLI	P24180 escherichia
5	656	32.5	385	1 YHUI_ECOLI	P37636 escherichia
6	432	21.4	415	1 YEGM_ECOLI	P76397 escherichia
7	297	14.7	371	1 MACA_YERPE	P58411 yersinia pe
8	286	14.2	382	1 Y894_HAEIN	O57500 haemophilus
9	269.5	13.3	371	1 MACA_ECOLI	P75830 escherichia
10	267.5	13.2	371	1 MACA_ECOLI	P58410 escherichia
11	241	11.9	520	1 CZCB_ALCEU	P13510 alcaligenes
12	236	11.7	520	1 CZCB_ALCEU	P94176 alcaligenes
13	217.5	10.8	397	1 NCCE_ALCEX	O44585 alcaligenes
14	194	9.6	395	1 CNRB_ALCEU	P37973 alcaligenes
15	187	9.3	430	1 SLUB_SALTY	O92480 salmonella
16	184	9.1	367	1 NOLF_RHIME	P25196 rhizobium m
17	183	9.1	328	1 Y873_YERPE	O93447 yersinia pe
18	178.5	8.8	343	1 YCCR_ECOLI	P32716 escherichia
19	173.5	8.6	332	1 YBHG_ECOLI	O6fjnc escherichia
20	172.5	8.5	332	1 YBHG_ECOLI	O6fjnc escherichia
21	171.5	8.5	332	1 YBHG_ECOLI	O6fjnc escherichia
22	169	8.4	285	1 YDHJ_ECOLI	P75777 escherichia
23	168.5	8.3	331	1 YBHG_ECOLI	P76185 escherichia
24	168.5	8.3	331	1 YBHG_ECOLI	O62879 salmonella
25	162	8.0	407	1 CUSB_SALTY	O82490 salmonella
26	156.5	7.8	390	1 E8RA_HAEIN	P77233 escherichia
27	154.5	7.7	310	1 YHCO_ECOLI	P44928 haemophilus
28	150.5	7.5	390	1 E8RA_ECOLI	P46482 escherichia
29	142.5	7.1	378	1 YIBH_ECOLI	P27303 escherichia
30	138.5	6.9	355	1 YHII_ECOLI	P32107 escherichia
31	136.5	6.8	334	1 YOH3_RHIME	P27626 escherichia
32	134.5	6.7	544	1 CH63_RHIME	O92444 rhizobium m
33	133.5	6.6	387	1 EWRK_ECOLI	O93090 rhizobium m
					P52599 escherichia

ALIGNMENTS

RESULT 1	ID	MTRC_NEIGO	STANDARD	PRT	412 AA
AC	P43505				
DT	01-NOV-1995 (Rel. 32, Created)				
DT	01-NOV-1995 (Rel. 32, Last sequence update)				
DT	15-JUL-1998 (Rel. 36, Last annotation update)				
DE	Membrane fusion protein mtrc precursor.				
GN	MTRC.				
OS	Neisseria gonorrhoeae.				
OC	Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;				
OC	Neisseriaceae; Neisseria.				
OX	NCBI_TaxID=485				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=FA19;				
RX	MEDLINE=94254732; PubMed=8196548;				
RA	Pan W., Spratt B.G.; PubMed=8196548;				
RY	"Regulation of the permeability of the gonococcal cell envelope by the mtr system."				
FL	Mol. Microbiol. 11:769-775(1994).				
CC	-1- FUNCTION: CELL MEMBRANE LIPOPROTEIN, INVOLVED IN CELL MEMBRANE PERMEABILITY TO HYDROPHOBIC COMPOUNDS SUCH AS ANTIBIOTICS, DYES AND DETERGENTS.				
CC	-1- SUBCELLULAR LOCATION: Attached to the inner membrane by a lipid anchor (potential).				
CC	-1- SIMILARITY: BELONGS TO THE ACRA/ACRE FAMILY.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).				
CC	EMBL, U14993; AAA80193.1; -				
DR	InterPro: IPR006143; H1YD.1.				
DR	Pfam: PF00529; H1YD.1.				
DR	PROSITE: PS00013; PROKAR LIPOPROTEIN. 1.				
KW	Signal; Inner membrane; Lipoprotein.				
FT	SIGNAL 1 24				
FT	POTENTIAL.				
FT	CHAIN 25 412				
FT	MEMBRANE FUSION PROTEIN MTRC.				
FT	LIPID 25 25				
FT	N-ACYL DIGLYCERIDE (POTENTIAL).				
SO	SEQUENCE 412 AA; 42773 MW; 97F9AFBCFAE321BA CRC64;				
Query Match	96.0%; Score 1938; DB 1; Length 412;				
Best Local Similarity	95.4%; Pred. No. 2.2e-107;				
Matches	393; Conservative 9; Mismatches 10; Indels 0; Gaps 0;				
OY	1 MAFYAFKAMRAALAAVALVLSGCGKGAAGGAGAEAPVGVVTPHQTVALT 60				
DB	1 MAFYAFKAMRAALAAVALVLSGCGKGAAGGAGAEAPVGVVTPHQTVALT 60				
OY	61 VELPGRLSRLTADVAOVGGIIQKRLFOEGSVYRAGOPLYOIDSSTYEANLSARAOLA 120				
DB	61 VELPGRLSRLTADVAOVGGIIQKRLFOEGSVYRAGOPLYOIDSSTYEANLSARAOLA 120				